HZ BOOKS
华章教育

国外电子与电气工程技术丛书

Springer

信号检测与估计

（原书第2版）

[美] H. Vincent Poor 著

廖桂生 杨志伟 等译

H. Vincent Poor

An Introduction to Signal Detection and Estimation

Second Edition

Springer

An Introduction to Signal Detection and Estimation

Second Edition



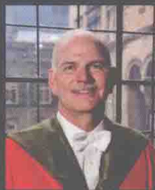
机械工业出版社
China Machine Press

本书是美国普林斯顿大学“统计信号处理”专业课教材，从信号检测概念、参数估计方法和波形估计理论三方面深入阐述信号检测与估计理论的基本思想，可作为高等院校通信类、信息类和控制类等专业高年级本科生或研究生的教材，也可供相关科研人员参考。

本书特色

- 解构与建构。无论是在离散域还是连续域，都能以简单实例解构抽象理论；贯穿全书的经典著作和文献建构起信号处理领域的知识网络，以科学发展的视角引导阅读之美。
- 循序与跨越。理论架构层层深入，部分证明和推导简捷清晰。充满挑战性的知识跨度要求读者具备应用概率论的基础知识，并启发读者独立思考，以欣赏作者的留白之美。
- 基石与桥梁。全书的公式和定理是一块块基石，搭起通向通信、自动控制等领域的桥梁，期待读者积极探索，在实践之中体验理论之美。

作者简介



H. Vincent Poor 美国普林斯顿大学工程与应用科学学院院长，美国国家科学院和工程院、英国皇家工程院、欧洲科学院院士，IEEE、IET会士，曾任IEEE《Transactions on Information Theory》杂志主编。Vincent Poor教授是信息与通信领域知名学者，在信息论、统计信号处理、随机理论等方向取得了突出的成果，并广泛应用于无线通信、社交网络以及智能电网中。Vincent Poor教授还是香港科技大学、爱丁堡大学等多所国际知名学校名誉博士，并获得过多项国际奖励和荣誉。



上架指导：信号处理


ISBN 978-7-111-48339-7

9 787111 483397 >

定价：49.00元

投稿热线：(010) 88379604
客服热线：(010) 88378991 88361066
购书热线：(010) 68326294 88379649 68995259

华章网站：www.hzbook.com
网上购书：www.china-pub.com
数字阅读：www.hzmedia.com.cn

封面设计·锡 彬

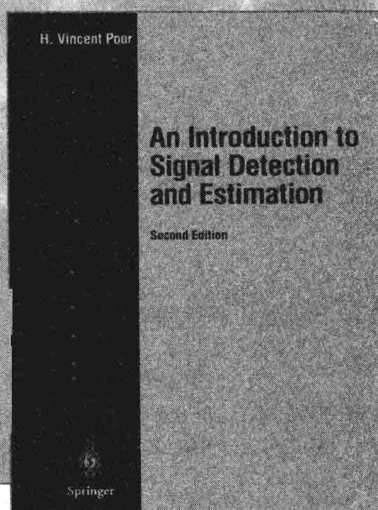
国外电子与电气工程技术丛书

信号检测与估计

（原书第2版）

[美] H. Vincent Poor 著
廖桂生 杨志伟 等译

An Introduction to Signal Detection and Estimation
Second Edition





机械工业出版社
China Machine Press

图书在版编目（CIP）数据

信号检测与估计（原书第 2 版）/（美）珀尔（Poor，H. V）著；廖桂生等译．—北京：机械工业出版社，2014.11
（国外电子与电气工程技术丛书）
书名原文：An Introduction to Signal Detection and Estimation, Second Edition

ISBN 978-7-111-48339-7

I. 信… II. ①珀… ②廖… III. ①信号检测 ②参数估计 IV. TN911.23

中国版本图书馆 CIP 数据核字（2014）第 244876 号

本书版权登记号：图字：01-2013-5973

本书介绍信号检测与估计的基本理论，主要包括三方面的内容。第 2、3、6 章讲述信号检测理论的基本概念，包括二元假设检验、复合假设检验、确知信号检测、部分参数确知信号检测、随机信号检测等内容。第 4 章讲述参数估计方法，包括贝叶斯参数估计方法、最大似然估计方法、向量参数估计、递归参数估计等内容。第 5、7 章讲述波形估计理论，包括 Kalman-Bucy 滤波、Wiener-Kolmogorov 滤波、线性/高斯估计、非线性滤波等内容。每章都有大量例题和习题，帮助读者理解抽象的理论知识。

本书可作为高等院校通信类、信息类和控制类等专业研究生或高年级本科生的专业教材或教学参考书，也可作为相关科研人员的参考用书。

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码：100037）

责任编辑：曲 熠	责任校对：殷 虹
印 刷：北京瑞德印刷有限公司	版 次：2015 年 1 月第 1 版第 1 次印刷
开 本：185mm×260mm 1/16	印 张：16
书 号：ISBN 978-7-111-48339-7	定 价：49.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换
客服热线：（010）88378991 88361066　　投稿热线：（010）88379604
购书热线：（010）68326294 88379649 68995259　　读者信箱：hzjsj@hzbook.com

译者序

本书英文版自出版以来，被全球众多著名大学选为研究生“统计信号处理”课程的教材或教学参考书。Poor 教授以多年的研究积淀和教学经验为基础撰写了本书，通过易于理解的方式，深入阐述信号检测和估计理论的基本思想。全书主要包括三方面的内容，第 2、3、6 章讲述信号检测理论的基本概念，包括离散域的二元假设检验、复合假设检验、序贯检测、非参数检测、稳健检测，连续域的确知信号检测、部分参数确知信号检测、随机信号检测等内容。第 4 章讲述参数估计方法，包括贝叶斯参数估计方法、最大似然估计方法、非随机参数估计的理论结构、向量参数估计、稳健参数估计、递归参数估计等内容。第 5、7 章讲述波形估计理论，包括离散域的 Kalman-Bucy 滤波、线性因果/非因果 Wiener-Kolmogorov 滤波，连续域的线性/高斯估计、非线性滤波等内容。本书最为突出的特点是提供了大量的实例，无论是在离散域还是连续域，都以简单实用的例子来描述检测与估计的基本理论与方法，并通过应用实例来比较各种方法的异同。此外，每章都有大量习题，这些习题与该章内容密切相关，很多习题是书中结论的证明或进一步阐述，以加深读者对抽象理论的理解。

本书要求读者具备理工科研究生基本课程中关于应用概率论及随机过程的相关知识。此外，在教学过程中，可根据需要进行内容取舍。

参加本书翻译工作的主要有廖桂生（前言，第 1、2、4 章）和杨志伟（第 3、5、6、7 章）。参加翻译、校对和译稿资料整理工作的还有刘志凌、郭小路、谢坚、党博、李东、粟嘉、何嘉懿、张学攀、许京伟、高永婵、杨东、束宇翔、杜文韬。全书由廖桂生进行审校。

由于译者水平有限，书中难免有不当之处，敬请读者批评指正。

前　言

本书旨在介绍信号检测与估计的基本理论。读者在阅读本书时，应当具备理工科研究生第一学期基本课程中关于应用概率论及随机过程的相关知识。这些预备知识可以参考 Wong（1983）的著作。信号检测与估计理论所涉及的更复杂的概念将主要在第 6 章和第 7 章中介绍，包括连续时间信号检测和估计问题。

本书改编自伊利诺伊大学和普林斯顿大学一个学期的第二阶段研究生课程。如果将授课范围限制在第 1～5 章，本书也可用于短期课程或者第一阶段课程，此时仅要求读者具备应用概率论的基本知识，包括随机向量和条件期望等概念。为了获取足够的背景知识，读者可以参考 Thomas（1986）的著作。

此外，还可以以其他教学模式使用本书。例如，第 2、3、6 章可用于介绍信号检测的短期课程，第 4、5、7 章可用于介绍信号估计的短期课程。类似地，还可以将第 1 章和第 4 章用于入门级课程，而将第 5 章和第 7 章用于高级课程。

在第 2 版的筹备过程中，听取了很多读者的意见和建议。由于人数众多，在此不便一一列出，但要对这些读者致以深深的谢意。

目录

第1章 引言

信号检测与估计是信号处理的一个研究领域，其目的是从信号中提取需要的信息。信号检测与估计的相关理论已在通信、自动控制等领域得到广泛应用。例如，在雷达和数据传输等通信领域，信号检测和估计是设计高效通信接收机的理论基础和分析工具。此外，在自动控制领域，检测与估计理论是精确推断进程状态和进行系统控制的基础。

雷达是检测与估计技术的一个应用实例。雷达的基本工作方式是发射一个电磁脉冲信号，然后等待目标可能反射的回波信号。受接收机噪声、大气扰动、地面和其他目标寄生反射以及信号失真等因素的影响，往往无法绝对准确地判断目标是否存在。因此，必须根据接收天线输出的(非理想的)观测结果来推断目标是否存在，此时，基于检测理论可设计出最优化的检测方法。进而，在判断出目标以某种概率存在之后，还希望估计目标的一些特性，比如位置和速度等，而这在广义上属于估计理论的范畴。上述估计结果可用于控制天线以跟踪目标或者远程控制目标以实现特定航迹飞行。检测估计技术的其他具体应用还包括地震学、射电天文学、声呐、语音和图像处理、医学信号处理及光通信等。

在检测与估计技术的实际应用中，通常会涉及基于观测值的推断处理，并且这些观测值往往存在不明原因导致的失真和缺损，甚至从这些观测值中提取的信息也是未知的。因此，将检测和估计问题置于概率框架下处理是非常有效的，此时可以将未知因素作为随机量来处理。就此而论，在统计推断领域内研究检测估计理论非常适合，并且在本书的论述过程中都将以这种方式进行阐述。

信号检测与估计理论学习中的一个基本概念是随机观测量 Y，它取值于观测值集合 Γ，其中 Γ 可以是向量的集合、波形的集合、实数的集合或者其他任意集合。我们期望从观测量 Y 中提取有关 Y 的一些特征信息。本书主要关注两类问题：检测问题和估计问题。在检测问题中，我们期望基于一些有限数量的可能情况或“自然状态”做出判决；在估计问题中，我们期望对一些无法直接观测的量进行估值。在上述两类问题中，观测值和待提取信息之间的关系均是概率性的，而不是直接必然的，在此意义上而言，Y 的统计特性受其自然状态或待估计量的真实值影响。因此，在建立检测/估计模型时，必须涉及基于 Γ 的一族概率分布，其中各元素均对应不同自然状态或在不同待估计量值下所呈现的统计状态。检测/估计模型一旦建立，则相应的检测/估计问题的目标即转化为找到一种处理观测量 Y 的最优方式，以便提取所需信息。区分这类问题的基本特征包括：待估计信息的性质(离散或连续)，关于待估计量或自然状态的先验信息的多少，以及用于比较不同检测估计过程的性能指标。

本书旨在向读者介绍检测与估计理论的基本原理。第 2 章、第 3 章和第 6 章主要讨论信号估计的相关问题。第 2 章介绍二元假设检验的基础知识，这部分知识是大多数信号检

测问题的基础。第 3 章将应用上述基础知识针对特定的信号检测模型推导最优检测器并分析其检测性能。此外，第 3 章还将讨论用于处理非标准情况下信号检测问题的几种特殊信号检测方法。第 2 章和第 3 章主要讨论观测值为向量时的情况，它与基于离散时间观测值(采样观测值)的信号检测问题相对应。基于连续时间观测值的信号检测问题将在第 6 章中进行讨论。连续时间模型的信号检测问题虽然与离散时间模型在本质上相同，但是由于对于这类问题的分析中涉及了更复杂的分析方法，因此将对其进行单独讨论。第 4 章、第 5 章和第 7 章讨论估计问题。第 4 章中给出了参数估计问题的基本原理和结构。与信号检测类似，由于分析难度不同，所以对离散时间估计(第 4 章、第 5 章)和连续时间估计(第 7 章)分别进行了讨论。

符号说明

为了详细描述观测集合 Γ 的概率分布，需要对 Γ 的子集进行概率分配。对于某些观测空间而言，无法为 Γ 的所有子集分配一致的概率值，因此，我们总是将 Γ 与它的一类子集 $\mathcal{G}$ 联系在一起，并期望对这类子集进行概率分配。$\mathcal{G}$ 中的集合称为*观测结果*，而组合(Γ, $\mathcal{G}$)则称为*观测空间*。为了便于分析，始终假设 $\mathcal{G}$ 满足 σ 代数，即假设 $\mathcal{G}$ 包含所有补集(关于 Γ)和其元素的可数并集。[⊖]

本书主要关注(Γ, $\mathcal{G}$)的两种情况。第一种情况是 $\Gamma=\mathbf{R}^n$，即 Γ 为 n 维实向量的集合；第二种情况是 Γ 为离散(可数)集合，且 $\Gamma=\{\gamma_1, \gamma_2, \cdots\}$。在第一种情况中，很自然地期望为形如 $\{y=(y_1, \cdots, y_n)^{\mathrm{T}}\in\mathbf{R}^n \mid a_1\leqslant y_1\leqslant b_1, \cdots, a_n\leqslant y_n\leqslant b_n\}$ 的集合分配概率，其中，a_i 和 b_i 均为任意实数。因此，对于 $\Gamma=\mathbf{R}^n$，把 $\mathcal{G}$ 看作包含 a_i 和 b_i 所确定的全部子集的最小 σ 代数，其中，a_i 和 b_i 在整个实数范围内取值。这类 σ 代数一般用 $\mathcal{B}^n$ 表示，称为 $\mathbf{R}^n$ 中的 Borel 集合。对于第二种情况，可以将 $\mathcal{G}$ 定义为 Γ 的所有子集。这类 σ 代数一般用 2^Γ 表示，称为 Γ 的幂集合。在上述两种观测空间内足以描述第 2 章～第 5 章中所讨论的大部分离散时间检测与估计问题。如无特别说明，均假设(Γ, $\mathcal{G}$)属于上述两种情况之一。在讨论连续时间的检测估计问题时会涉及更为抽象的观测空间，相关内容会在第 6 章和第 7 章加以介绍。

对于离散观测空间(Γ, 2^Γ)，以*概率质量函数*的形式为 Γ 中的子集分配概率，即 p: $\Gamma\to[0, 1]$，且

$$P(A)=\sum_{\gamma_i\in A}p(\gamma_i),\quad A\in 2^\Gamma \tag{1.1}$$

其中，$P(A)$表示观测量 Y 的取值位于集合 A 中的概率。如果满足条件 $\sum_{i=1}^{\infty}p(\gamma_i)=1$，则将 Γ 映射到$[0, 1]$的任意函数均是合理的概率质量函数。对于观测空间($\mathbf{R}^n$, $\mathcal{B}^n$)，则主要关注所谓的*连续随机向量*，且可按照*概率密度函数*的形式对其进行概率分配，即 p: $\mathbf{R}^n\to[0, \infty)$，且[注意：式(1.2)中的积分为 n 重积分]

$$P(A)=\int_A p(y)\mathrm{d}y,\quad A\in\mathcal{B}^n \tag{1.2}$$

⊖ 也就是说，$\mathcal{G}$ 具有下述性质：若 $A\in\mathcal{G}$，则有 $A^c\in\mathcal{G}$(本书中上标 c 表示取补集)；若 $A_1, A_2, \cdots\in\mathcal{G}$，则有 $\bigcup_{i=1}^{\infty}A_i\in\mathcal{G}$。

如果满足条件$\int_{\mathbf{R}^n} p(y)\mathrm{d}y = 1$，则将$\mathbf{R}^n$映射到$[0, \infty)$的任意可积函数均是合理的概率密度函数。为了简化术语和公式表述，对概率质量函数和概率密度函数均使用密度一词，且将式(1.1)中的求和公式和式(1.2)中的积分公式统一用下述公式表示

$$P(A) = \int_A p(y)\mu(\mathrm{d}y) \tag{1.3}$$

其中，也可以省略积分变量，将上式进一步简写为

$$P(A) = \int_A p\,\mathrm{d}\mu \tag{1.4}$$

对于以随机观测量Y为自变量的实值函数g，经常需要求解$g(Y)$的期望值，即$E\{g(Y)\}$。对于离散观测空间$(\Gamma, 2^{\Gamma})$，上述期望值为

$$E\{g(Y)\} = \sum_{i=1}^{\infty} g(\gamma_i)p(\gamma_i) \tag{1.5}$$

而对于$(\mathbf{R}^n, \mathcal{B}^n)$中的连续随机向量，有

$$E\{g(Y)\} = \int_{\mathbf{R}^n} g(y)p(y)\mathrm{d}y \tag{1.6}$$

其中，对于每一种情况，均假设相应的求和或积分是存在的。同样地，为了简化公式，采用下述公式对式(1.5)和式(1.6)进行统一表述

$$E\{g(Y)\} = \int_{\Gamma} g(y)p(y)\mu(\mathrm{d}y) = \int_{\Gamma} gp\,\mathrm{d}\mu \tag{1.7}$$

对上述公式的进一步阐述将在第6章中给出。易知，式(1.3)和式(1.4)是式(1.7)中的函数g按下式定义时的特殊情况

$$g(y) = \begin{cases} 1, & y \in A \\ 0, & y \in A^c \end{cases} \tag{1.8}$$

在后续的讨论中，均采用大写字母表示随机量，并采用小写字母表示这些随机量的特定取值。因此，本书的表述方式为：随机观测量Y的可能取值为y。

第2章

假设检验基础

2.1 引言

大多数信号检测问题都可以归结为 M 元假设检验问题，这一假设以观测数据(可用向量或函数表示)为基础，通过判决 M 元统计信号来描述观测数据。例如，在 M 元通信接收系统中，观测数据的波形为 M 元信号之一，该信号会受到随机信道或接收机噪声的干扰，因此我们期望通过统计判决来确定可表示观测数据的具体信号形式。显然，对于任意给定的判决问题，都有一系列可行的判决策略或准则可以应用，然而，本书希望选择一种在某种意义上是最优的判决准则。尽管就最优性而言，存在多种定义，但本章仅考虑三种最常见的准则——贝叶斯准则、极小化极大准则和尼曼-皮尔逊准则，并推导这些准则的最优解。为了不失一般性，虽然本章中的大多数结论可直接推广到 M 元假设检验(在习题中会具体探讨)，但是本章只考虑二元假设检验问题。第 3 章和第 6 章将详细讨论这些理论在信号检测中的具体应用。

2.2 贝叶斯假设检验

本章所考虑的首要问题是简单的二元假设检验问题，它有两种可能的假设，称为“状态或类型”，即 H_0 和 H_1，分别对应观测空间$(\Gamma, \mathcal{G})$中的两种可能的概率分布 P_0 和 P_1。可将问题描述如下

$$
\begin{aligned}
H_0 &: Y \sim P_0 \\
H_1 &: Y \sim P_1
\end{aligned}
\tag{2.2.1}
$$

式中，记号“$Y \sim P$”表示“Y 满足 P 分布”。假设 H_0 和假设 H_1 有时也称为零假设和非零假设。关于假设 H_0 和假设 H_1 的判决准则 δ 将观测集 Γ 分成 $\Gamma_1 \in \mathcal{G}$和 $\Gamma_0 = \Gamma_1^c$两个子集，如果 $j=0$ 或 1 时存在 $y \in \Gamma_j$，则判决假设 H_j 成立。集合 Γ_1 称为拒绝域(或关键域)，集合 Γ_0 称为接受域。也可将判决准则 δ 看成是由下式给出的在集合 Γ 上的函数，即

$$
\delta(y) = \begin{cases} 1, & y \in \Gamma_1 \\ 0, & y \in \Gamma_1^c \end{cases}
\tag{2.2.2}
$$

对给定的 $y \in \Gamma$，δ 的取值等于判决规则 δ 所接受假设的下标值。

本章旨在以某种最优的方法选择 Γ_1，在这种思路的指引下，需要对不同判决赋予不同的代价。特别地，假设正数 $C_{ij}(i=0,1,\ j=0,1)$是在假设 H_j 为真的情况下，判决假设 H_i 成立所需付出的代价。对于每个假设，可将条件风险定义为：当该假设为真时，由判决准则所引发的平均代价或期望代价，即

$$
R_j(\delta) = C_{1j} P_j(\Gamma_1) + C_{0j} P_j(\Gamma_0), \quad j = 0,1
\tag{2.2.3}
$$

注意，$R_j(\delta)$是当 H_j 为真时选择 H_1 的代价与选择 H_1 的概率的乘积加上选择 H_0 的代价与选择 H_0 的概率的乘积。

假设 H_0 和 H_1 的概率是 π_0 和 $\pi_1=1-\pi_0$，即 π_j 是假设 H_j 为真的概率，它与 Y 的取值无关。概率 π_0 和 π_1 也称为两种假设的先验概率。给定先验概率后，平均风险或贝叶斯风险可用来描述判决准则 δ 的总平均代价，表达式由下式给出

$$r(\delta)=\pi_0 R_0(\delta)+\pi_1 R_1(\delta) \tag{2.2.4}$$

根据贝叶斯风险的定义，可设计关于假设 H_0 和假设 H_1 的最优判决准则，该判决准则在所有判决准则中具有最小的贝叶斯风险，称为关于 H_0 和 H_1 的贝叶斯准则。

综合式(2.2.3)和式(2.2.4)并利用 $P_j(\Gamma_1^c)=1-P_j(\Gamma_1)$后，可以得到

$$\begin{aligned} r(\delta) &= \sum_{j=0}^{1}\pi_j\left[C_{0j}(1-P_j(\Gamma_1))+C_{1j}P_j(\Gamma_1)\right] \\ &= \sum_{j=0}^{1}\pi_j C_{0j}+\sum_{j=0}^{1}\pi_j(C_{1j}-C_{0j})P_j(\Gamma_1) \end{aligned} \tag{2.2.5}$$

假设 P_j 的概率密度是 $p_j(j=0，1)$，根据第 1 章的符号说明，可将式(2.2.5)表示为

$$r(\delta)=\sum_{j=0}^{1}\pi_j C_{0j}+\int_{\Gamma_1}\left[\sum_{j=0}^{1}\pi_j(C_{1j}-C_{0j})p_j(y)\right]\mu(\mathrm{d}y) \tag{2.2.6}$$

如果按照下式选取 Γ_1，则 $r(\delta)$可取得极小值，即

$$\begin{aligned} \Gamma_1 &= \left\{y\in\Gamma \,\middle|\, \sum_{j=0}^{1}\pi_j(C_{1j}-C_{0j})p_j(y)\leqslant 0\right\} \\ &= \{y\in\Gamma \mid \pi_1(C_{11}-C_{01})p_1(y)\leqslant\pi_0(C_{00}-C_{10})p_0(y)\} \end{aligned} \tag{2.2.7}$$

假设 $C_{11}<C_{01}$(正确选择 H_1 的代价小于错误拒绝 H_1 的代价)，式(2.2.7)可以重写成

$$\Gamma_1=\{y\in\Gamma \mid p_1(y)\geqslant\tau p_0(y)\} \tag{2.2.8}$$

式中

$$\tau\triangleq\frac{\pi_0(C_{10}-C_{00})}{\pi_1(C_{01}-C_{11})} \tag{2.2.9}$$

注意，区域$\{y\in\Gamma \mid p_1(y)=\tau p_0(y)\}$对平均误差没有影响，因此可以根据需要将其从 Γ_1 中全部或部分省略。

将式(2.2.8)中由拒绝域所描述的判决准则定义为似然比检验(或概率比检验)，该检验在假设检验理论中处于核心地位。注意，式(2.2.8)中的 Γ_1 可以重写成

$$\Gamma_1=\{y\in\Gamma \mid [p_1(y)/p_0(y)]\geqslant\tau\} \tag{2.2.10}$$

其中，定义对于任何 $k\geqslant0$ 的情况，$k/0$ 的值为∞。H_0 和 H_1 之间的似然比定义为

$$L(y)=\frac{p_1(y)}{p_0(y)},\quad y\in\Gamma \tag{2.2.11}$$

将上式定义为 H_0 和 H_1 之间的似然比(或似然比统计)。显然，式(2.2.8)所描述的贝叶斯判决准则首先需要计算出基于 Y 的观测值的似然比，然后通过与门限值 τ 作对比来完成判决。换言之，贝叶斯判决准则可写成

$$\delta_{\mathrm{B}}(y)=\begin{cases}1, & L(y)\geqslant\tau\\ 0, & L(y)<\tau\end{cases} \tag{2.2.12}$$

此外，均匀代价是常用的代价分配方法，即

$$C_{ij}=\begin{cases}0, & i=j\\1, & i\neq j\end{cases} \tag{2.2.13}$$

与之对应的贝叶斯风险为

$$r(\delta)=\pi_0 P_0(\Gamma_1)+\pi_1 P_1(\Gamma_0) \tag{2.2.14}$$

注意，$P_i(\Gamma_j)$是当H_i为真时，选择H_j的概率。因此，对于$i\neq j$的情况，$P_i(\Gamma_j)$是H_i为真时发生误判的概率，$r(\delta)$是由判决准则δ产生的平均错误概率。又因为似然比判决$\tau=\pi_0/\pi_1$可最小化式(2.2.13)对应的贝叶斯风险$r(\delta)$，所以它也是最小错误概率判决。

根据贝叶斯准则[参见 Thomas(1986)]，在随机观测值Y取值为y的条件下，H_j为真的条件概率可表示为

$$\pi_j(y)=P(H_j\text{ 为真}\mid Y=y)=\frac{p_j(y)\pi_j}{p(y)} \tag{2.2.15}$$

式中，$p(y)=\pi_0 p_0(y)+\pi_1 p_1(y)$表示$Y$的平均或全局概率密度。$\pi_0(y)$和$\pi_1(y)$称为两种假设的后验概率。利用式(2.2.15)，式(2.2.7)所描述的贝叶斯判决准则的关键域可重写成

$$\Gamma_1=\{y\in\Gamma\mid C_{10}\pi_0(y)+C_{11}\pi_1(y)\leqslant C_{00}\pi_0(y)+C_{01}\pi_1(y)\} \tag{2.2.16}$$

因此，在贝叶斯判决准则框架中，最优判决是基于后验概率的。即利用观测值将先验概率更新为后验概率。与之对应的后验代价为

$$C_{i0}\pi_0(y)+C_{i1}\pi_1(y) \tag{2.2.17}$$

式(2.2.17)是已知Y等于y时，选择假设H_i所产生的平均代价。因此贝叶斯准则通过选择具有最小后验代价的假设来实现判决。例如，对于由式(2.2.13)描述的均匀代价准则而言，贝叶斯准则可以写成

$$\delta_B(y)=\begin{cases}1, & \pi_1(y)\geqslant\pi_0(y)\\0, & \pi_1(y)<\pi_0(y)\end{cases} \tag{2.2.18}$$

因此，最小错误概率判决准则通过选取$Y=y$时具有最大后验概率的假设实现判决。该判决准则也称为二元假设检验的 MAP(maximum a posteriori)判决。

下面的例子将详细阐明贝叶斯判决准则。

例 2.2.1　二元信道

假设通信信道内数据以二进制(即“0”或“1”)进行传输，观测值Y的取值为0或1。由于存在信道噪声、非理想调制或解调等原因，信道中传输的“0”可能以概率λ_0被接收成“1”，以概率$1-\lambda_0$被接收成“0”，其中$0<\lambda_0<1$。类似地，传输的“1”以概率λ_1被接收成“0”，以概率$1-\lambda_1$被接收成“1”，其中$0<\lambda_1<1$(具体描述见图2.2.1)。因此，由观测值Y无法精确判定传输的数字究竟是“0”还是“1”，需要寻求某个优化准则以实现判决。

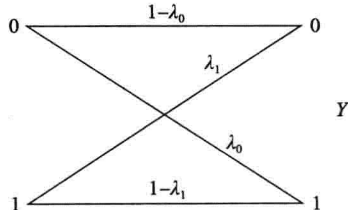


图 2.2.1　二元信道

该问题可建模成二元假设检验问题，其中，假设H_j表示“j”被传输($j=0, 1$)，观测集Γ是$\{0, 1\}$，观测值Y的概率密度可表示为

$$p_j(y)=\begin{cases}\lambda_j, & y\neq j\\1-\lambda_j, & y=j\end{cases} \tag{2.2.19}$$

其中，$j=0$ 和 1。因此，似然比可由下式给出

$$L(y)=\frac{p_1(y)}{p_0(y)}=\begin{cases}\dfrac{\lambda_1}{1-\lambda_0}, & y=0\\ \dfrac{1-\lambda_1}{\lambda_0}, & y=1\end{cases} \tag{2.2.20}$$

对于贝叶斯检验，根据式(2.2.9)中的代价和先验概率来确定门限值 τ。如果 λ_1、λ_0 和 τ 满足 $\lambda_1 \geqslant \tau(1-\lambda_0)$，式(2.2.12)所表示的似然比检验可解释为将接收的"0"判决为传输的"1"；否则，接收的"0"对应传输的"0"。类似地，如果 $1-\lambda_1 \geqslant \tau\lambda_0$，似然比检验就可解释为将传输的"1"接收成"1"；如果 $1-\lambda_1 < \tau\lambda_0$，则接收到的"1"将被看成是"0"。对于本例，最小的贝叶斯风险 $r(\delta_B)$ 可以由式(2.2.5)直接计算出来(参见习题 1)。

例如，对于均匀代价且先验概率相等($\pi_0=\pi_1=1/2$)的情况，有 $\tau=1$，且式(2.2.12)的贝叶斯准则变为

$$\delta_B(0)=\begin{cases}1, & \lambda_1 \geqslant (1-\lambda_0)\\ 0, & \lambda_1 < (1-\lambda_0)\end{cases} \tag{2.2.21a}$$

$$\delta_B(1)=\begin{cases}1, & (1-\lambda_1) \geqslant \lambda_0\\ 0, & (1-\lambda_1) < \lambda_0\end{cases} \tag{2.2.21b}$$

根据前面的讨论，$L(y)$ 等于 τ 的边界点可任意判决为 Γ_1 或者 Γ_0。因此从风险角度出发，等价的贝叶斯检验可表示为

$$\delta_B(y)=\begin{cases}y, & (1-\lambda_1) \geqslant \lambda_0\\ 1-y, & (1-\lambda_1) < \lambda_0\end{cases} \tag{2.2.22}$$

如果进一步假设信道对称，即 $\lambda_1=\lambda_0=1$，则式(2.2.22)变成

$$\delta_B(y)=\begin{cases}y, & \lambda \leqslant 1/2\\ 1-y, & \lambda > 1/2\end{cases} \tag{2.2.23}$$

式(2.2.23)的含义非常直观：如果信道发生比特翻转的概率较高($\lambda>1/2$)，则接收端需要进行比特翻转，即通过翻转所接收到的比特来作判决；反之，接收端不进行比特翻转。对于后一种情况，最小贝叶斯风险为

$$r(\delta_B)=\min\{\lambda, 1-\lambda\} \tag{2.2.24}$$

因此，直接传输或翻转比特后再传输的信道越可靠，传输性能就越好。值得注意的是，在均匀代价和相等的先验概率条件下，即使缺乏观测值 y，随机猜测的风险也是 1/2。因此，就 $\lambda=1/2$ 而言，观测是没有价值的。◀

例 2.2.2　高斯误差下的位置检验

考虑关于实值观测值 Y 的两种假设

$$\begin{aligned}H_0: Y &= \varepsilon+\mu_0\\ H_1: Y &= \varepsilon+\mu_1\end{aligned} \tag{2.2.25}$$

其中，ε 是均值为 0、方差为 σ^2 的高斯随机变量，μ_0 和 μ_1 是两个固定值，并满足 $\mu_1 > \mu_0$。注意，由于 μ_0 或 μ_1 与 ε 的和仅改变观测值的均值，因此需要对观测数据的两种分布或"位

置”进行判决。从观测空间的分布角度看，式(2.2.25)的两个假设可重写成

$$\begin{aligned} H_0: Y &\sim \mathcal{N}(\mu_0, \sigma^2) \\ H_1: Y &\sim \mathcal{N}(\mu_1, \sigma^2) \end{aligned} \tag{2.2.26}$$

其中，$\mathcal{N}(\mu, \sigma^2)$表示均值为$\mu$、方差为$\sigma^2$的高斯(正态)分布。服从$\mathcal{N}(\mu, \sigma^2)$分布的随机变量的概率密度可表示为

$$(1/\sqrt{2\pi}\sigma)\exp\{-(x-\mu)^2/2\sigma^2\}, \quad x \in \mathbf{R}$$

式(2.2.26)的似然比为

$$L(y) = \frac{p_1(y)}{p_0(y)} = \frac{\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(y-\mu_1)^2/2\sigma^2}}{\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(y-\mu_0)^2/2\sigma^2}} = \exp\left\{\frac{\mu_1-\mu_0}{\sigma^2}\left(y-\frac{\mu_0+\mu_1}{2}\right)\right\} \tag{2.2.27}$$

即式(2.2.26)的贝叶斯检验为

$$\delta_{\mathrm{B}}(y) = \begin{cases} 1, & \exp\left\{\dfrac{\mu_1-\mu_0}{\sigma^2}\left(y-\dfrac{\mu_0+\mu_1}{2}\right)\right\} \geqslant \tau \\ 0, & \text{其他} \end{cases} \tag{2.2.28}$$

其中，τ是判决门限。因为$\mu_1 > \mu_0$，所以式(2.2.27)的似然比是关于观测值y的严格增函数$[\mathrm{d}L(y)/\mathrm{d}y=(\mu_1-\mu_0)L(y)/\sigma^2>0]$。所以，将$L(y)$与门限值$\tau$作比较等价于将$y$与另一个门限值$\tau'=L^{-1}(\tau)$作比较，其中$L^{-1}$是$L$的反函数。特别地，对式(2.2.28)中不等式分别取对数，并重新整理之后得到

$$\delta_{\mathrm{B}}(y) = \begin{cases} 1, & y \geqslant \tau' \\ 0, & y < \tau' \end{cases} \tag{2.2.29}$$

其中

$$\tau' = \frac{\sigma^2}{\mu_1-\mu_0}\log(\tau) + \frac{\mu_0+\mu_1}{2} \tag{2.2.30}$$

例如，在均匀代价和先验概率相等的情况下，可以得到$\tau=1$和$\tau'=(\mu_0+\mu_1)/2$。此时，贝叶斯准则将观测值与μ_0和μ_1的平均值作比较。如果y大于或等于平均值，选择H_1；如果y小于平均值，选择H_0。图2.2.2是与之对应的判决示意图。

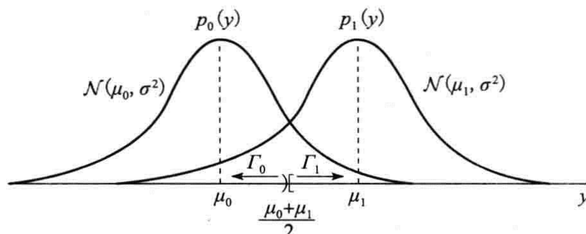


图 2.2.2　高斯误差、均匀代价和等先验概率下的位置检验

如果已知$P_j(\Gamma_1)(j=0, 1)$，则最小贝叶斯风险$r(\delta_{\mathrm{B}})$可由式(2.2.5)计算得到。又因为$\Gamma_1=\{y\in\mathbf{R} \mid y\geqslant\tau'\}$，进而可以得到

$$P_j(\Gamma_1) = \int_{\tau'}^{\infty} p_j(y)\mathrm{d}y = 1-\Phi\left(\frac{\tau'-\mu_j}{\sigma}\right) = \begin{cases} 1-\Phi\left(\dfrac{\log\tau}{d}+\dfrac{d}{2}\right), & j=0 \\ 1-\Phi\left(\dfrac{\log\tau}{d}-\dfrac{d}{2}\right), & j=1 \end{cases} \tag{2.2.31}$$

其中，Φ表示满足$\mathcal{N}(0, 1)$分布的随机变量的累积概率分布函数，$d \triangleq (\mu_1-\mu_0)/\sigma$。对于均匀代价和先验概率相等的特殊情况，可以直接得到

$$r(\delta_B) = 1 - \Phi(d/2) \tag{2.2.32}$$

图 2.2.3 给出了贝叶斯风险与 d 的关系。注意，风险随均值$(\mu_1 - \mu_0)$与误差标准差 σ 的比值的增大而单调递减。d 可理解为信噪比，在接下来的章节中将会深入介绍其含义。◀

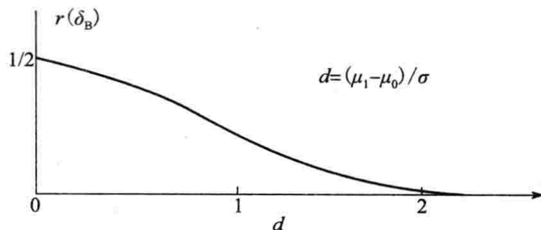


图 2.2.3　高斯误差下位置判决的贝叶斯风险

虽然例 2.2.1 和例 2.2.2 相当简单，但却说明了贝叶斯假设检验的基本原理。第 3 章和第 4 章将讨论更复杂的例子，本章末还有一系列相关习题。

本节最重要的结论是：给定代价 C_{ij} 和先验概率 π_i 后，式(2.2.12)所描述的似然比是贝叶斯意义上的最优判决准则。下一节将讨论在先验概率与(或)代价未知情况下的其他最优判决准则。

2.3　极小化极大假设检验

本节考虑先验概率 π_0 和 π_1 未知的判决准则设计问题，这是因为在设计判决准则时通常无法控制或精确知道系统的运行机制。此时，最小化平均或贝叶斯风险都不是可接受的设计准则，因为不太可能利用单一的判决准则来最小化每个可能的先验概率分布的平均风险。因此，有必要寻求新的判决准则，该准则相对所有的判决准则 δ 而言能使最大的条件风险 $R_0(\delta)$和 $R_1(\delta)$最小化，即判决准则最小化性能指标

$$\max\{R_0(\delta), R_1(\delta)\} \tag{2.3.1}$$

该判决准则也称为极小化极大准则。接下来的章节将会讨论该判决准则的具体实现方法。

为了寻求最小化式(2.3.1)的判决准则，首先考虑已知先验概率 $\pi_0 \in [0, 1]$和判决准则 δ 的平均风险 $r(\pi_0, \delta)$，即

$$r(\pi_0, \delta) = \pi_0 R_0(\delta) + (1 - \pi_0) R_1(\delta) \tag{2.3.2}$$

在图 2.3.1 中，π_0 的函数 $r(\pi_0, \delta)$是一条从 $r(0, \delta) = R_1(\delta)$到 $r(1, \delta) = R_0(\delta)$的直线。因此，当 δ 值固定，且 π_0 在 0～1 之间取值时，$r(\pi_0, \delta)$的最大值不是在 $\pi_0 = 0$ 处，就是在 $\pi_0 = 1$ 处，并且最大值是 $\max\{R_0(\delta), R_1(\delta)\}$。所以最小化式(2.3.1)的问题等价于在整个 δ 上最小化下式

$$\max_{0 \leqslant \pi_0 \leqslant 1} r(\pi_0, \delta) \tag{2.3.3}$$

相比式(2.3.1)而言，式(2.3.3)更容易求解。

对于每个先验概率 $\pi_0 \in [0, 1]$，假设 δ_{π_0} 表示对应该先验概率的贝叶斯判决准则。令 $V(\pi_0) = r(\pi_0, \delta_{\pi_0})$表示关于先验概率 π_0 的最小贝叶斯风险。易知 $V(\pi_0)$是关于 π_0 的连续凹函数，其中 $\pi_0 \in [0, 1]$，$V(0) = C_{11}$，$V(1) = C_{00}$(参见习题 8)。图 2.3.1 画出了典型的 $V(\pi_0)$曲线。

为了讨论方便，假设 $V(\pi_0)$和 $r(\pi_0, \delta)$如图 2.3.1 所示。其中还有一条标有 $r(\pi_0, \delta_{\pi_0'})$的直线，该直线平行于 $r(\pi_0, \delta)$，并且相切于 $V(\pi_0)$。注意，对于这种情况，δ 不是极小化极大准则，因为 $r(\pi_0, \delta_{\pi_0'})$所表示的代价直线全部在直线 $r(\pi_0, \delta)$之下，即 $r(\pi_0, \delta_{\pi_0'})$的最大值小于 $r(\pi_0, \delta)$。又因为 $r(\pi_0, \delta_{\pi_0'})$和 $V(\pi_0)$在点 $\pi_0 = \pi_0'$ 处相切，所以 $\delta_{\pi_0'}$ 是给

定先验概率 π'_0 的贝叶斯准则。显然，对于任意判决方法 δ，都可以画出类似的正切线(低于两个条件风险)，因此只有贝叶斯准则才有可能是极小化极大准则。此外，通过分析图 2.3.2，在 V 有内部最大值时，极小化极大准则就是在 $\pi_0\in[0, 1]$上最大化 $V(\pi_0)$的先验概率 π_L 所对应的贝叶斯准则。注意，对于这个先验概率，易知 $r(\pi_0, \delta_{\pi_L})$是关于 π_0 的常量，即 $\max\{R_0(\delta_{\pi_L}), R_1(\delta_{\pi_L})\}=R_0(\delta_{\pi_L})=R_1(\delta_{\pi_L})$(条件风险相等的判决方法称为均衡准则)。事实上，从图 2.3.2 易知，δ_{π_L} 是极小化极大值。因为如果 $\pi'_0<\pi_L$，有 $\max\{R_0(\delta_{\pi'_0}), R_1(\delta_{\pi'_0})\}=R_0(\delta_{\pi'_0})>R_0(\delta_{\pi_L})$；如果 $\pi''_0>\pi_L$，有 $\max\{R_0(\delta_{\pi''_0}), R_1(\delta_{\pi''_0})\}=R_1(\delta_{\pi''_0})>R_1(\delta_{\pi_L})$。

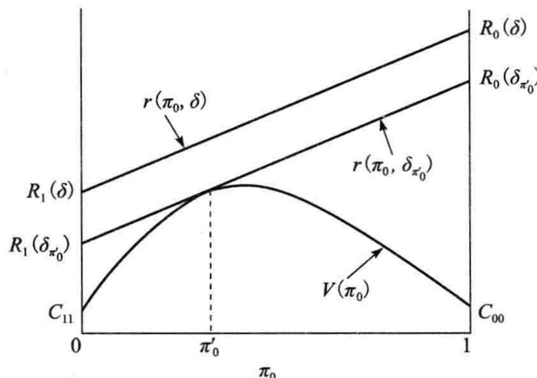


图 2.3.1 函数 $r(\pi_0, \delta)$和 $V(\pi_0)$

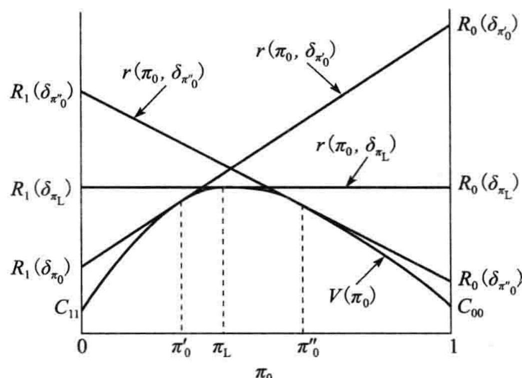


图 2.3.2 当 V 有内部最大值时的极小化极大值

图 2.3.2 中，π_L 对应极大化极小贝叶斯风险，因此也称为最不利先验概率。此时，极小化极大判决方法是关于最不利先验概率的贝叶斯准则。虽然在以上讨论中，没有考虑 $\max V(\pi_0)(0\leqslant\pi_0\leqslant 1)$可能会出现在 $\pi_0=0$ 或 $\pi_0=1$ 的情况，也不能保证 $V(\pi_0)$在每一点上都有直线与之相切(即该函数不是在所有点上可微)。但是，极小化极大准则几乎总是关于最不利先验概率的贝叶斯准则。下面推导极小化极大假设检验问题的一般解。首先讨论如下命题，其中 $V(\pi_0)$如图 2.3.1 和图 2.3.2 所示，即 $\pi_L=0$ 或 $\pi_L=1$。

命题 2.3.1 极小化极大检验

假设 π_L 是式 $V(\pi_L)=\max V(\pi_0)(0\leqslant\pi_0\leqslant 1)$的解，假设 $\pi_L=0$ 或 $\pi_L=1$ 或 $R_1(\delta_{\pi_L})=R_0(\delta_{\pi_L})$，则 δ_{π_L} 是极小化极大准则。

证明 首先，考虑 $R_1(\delta_{\pi_L})=R_0(\delta_{\pi_L})$的情况。对于任何先验概率 π_0，有

$$\max_{0\leqslant\pi_0\leqslant 1}\min_{\delta} r(\pi_0,\delta) = r(\pi_L,\delta_{\pi_L}) = r(\pi_0,\delta_{\pi_L}) \tag{2.3.4}$$

其中，第一个等式利用了 V 和 π_L 的定义，第二个等式利用了 $r(\pi_0, \delta_{\pi_L})$在 π_0 处是常数这一事实。进而可以得到

$$\max_{0\leqslant\pi_0\leqslant 1}\min_{\delta} r(\pi_0,\delta) = \max_{0\leqslant\pi_0\leqslant 1} r(\pi_0,\delta_{\pi_L}) \geqslant \min_{\delta}\max_{0\leqslant\pi_0\leqslant 1} r(\pi_0,\delta) \tag{2.3.5}$$

对于每个 δ 而言，都有

$$\max_{0\leqslant\pi_0\leqslant 1} r(\pi_0,\delta) \geqslant \max_{0\leqslant\pi_0\leqslant 1}\min_{\delta} r(\pi_0,\delta) \tag{2.3.6}$$

这意味着

$$\min_{\delta}\max_{0\leqslant\pi_0\leqslant 1} r(\pi_0,\delta) \geqslant \max_{0\leqslant\pi_0\leqslant 1}\min_{\delta} r(\pi_0,\delta) \tag{2.3.7}$$

联合式(2.3.5)和式(2.3.7)，可以得到

$$\min_{\delta}\max_{0\leqslant\pi_0\leqslant 1} r(\pi_0,\delta)=\max_{0\leqslant\pi_0\leqslant 1}\min_{\delta} r(\pi_0,\delta) \tag{2.3.8}$$

此时，式(2.3.4)左边的等式等价为

$$r(\pi_L,\delta_{\pi_L})=\min_{\delta}\max_{0\leqslant\pi_0\leqslant 1} r(\pi_0,\delta) \tag{2.3.9}$$

即 δ_{π_L} 是极小化极大准则。 □

现在假设 $\pi_L=0$，易知 $\max r(\pi_0,\delta_{\pi_L})=R_1(\delta_{\pi_L})=r(\pi_L,\delta_{\pi_L})(0\leqslant\pi_0\leqslant 1)$。在图 2.3.3 中给出了这种典型情况。结合式(2.3.6)～式(2.3.9)，易知 δ_{π_L} 同样是极小化极大准则。类似的结论在 $\pi_L=1$ 的情况下也成立。

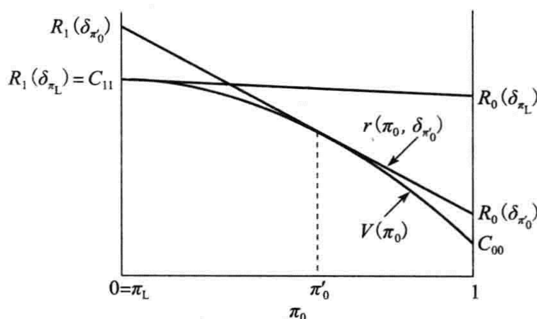


图 2.3.3　端点处取最大值的极小化极大准则

注意，对于任意的 $\pi'_0\in[0,1]$和 $\pi_0\in[0,1]$，都有 $r(\pi_0,\delta_{\pi'_0})\geqslant V(\pi_0)$和 $r(\pi'_0,\delta_{\pi'_0})\geqslant V(\pi'_0)$成立。如图 2.3.2 所示，既然 $r(\pi_0,\delta_{\pi'_0})$是关于 π_0 的直线，那么它一定是 V 在点 $\pi_0=\pi'_0$ 处的切线。

因此，如果 V 在 π'处可微，一定可以推出

$$V'(\pi'_0)=\mathrm{d}r(\pi_0,\delta_{\pi'_0})/\mathrm{d}\pi_0=[R_0(\delta_{\pi'_0})-R_1(\delta_{\pi'_0})]$$

其中，V'表示 V 的导数。如果 V 的最大值不在区间的端点，即 $\pi_L\in(0,1)$，则 V 在 π_L 处可微且必有 $V'(\pi_L)$等于 0。(注意，V 是凹函数，其导数 $V'(\pi_0)$在最大值处等于 0。)所以，不论 $\pi_L\in(0,1)$或 $V'(\pi_L)$是否存在，条件 $R_1(\delta_{\pi_L})=R_0(\delta_{\pi_L})$都成立。因此，命题 2.3.1 的假设是非强制性的。

至此，已经给出了除最大值处不可微这一情况之外的所有情况下的极小化极大准则的特性。对于最大值处不可微的情况，需要定义 $\delta^-_{\pi_L}=\lim_{\pi_0\uparrow\pi_L}\delta_{\pi_0}$ 和 $\delta^+_{\pi_L}=\lim_{\pi_0\downarrow\pi_L}\delta_{\pi_0}$ 两种判决准则。值得注意的是，无论贝叶斯判决准则 δ_{π_0} 采用何种定义，$\delta^-_{\pi_L}$ 的关键域都可表示为

$$\Gamma_1^-=\{y\in\Gamma\,|\,(1-\pi_L)(C_{11}-C_{01})p_1(y)\leqslant\pi_L(C_{00}-C_{10})p_0(y)\} \tag{2.3.10}$$

类似地，$\delta^+_{\pi_L}$ 的关键域为(假设 $C_{11}<C_{01}$)

$$\Gamma_1^+=\{y\in\Gamma\,|\,(1-\pi_L)(C_{11}-C_{01})p_1(y)<\pi_L(C_{00}-C_{10})p_0(y)\} \tag{2.3.11}$$

对于 $q\in[0,1]$，考虑判决准则 $\tilde{\delta}_{\pi_L}$ 选择关键域 Γ_1^- 的概率为 q，Γ_1^+ 的概率为$(1-q)$，即如果 $y\in\Gamma_1^+$，$\tilde{\delta}_{\pi_L}$ 选择 H_1；如果 $y\in(\Gamma_1^-)^c$，$\tilde{\delta}_{\pi_L}$ 选择 H_0；如果 y 在 Γ_1^- 的边界上，以概率 q 选择 H_1。从式(2.2.6)可知，在边界上的判决与贝叶斯风险无关，即 $r(\pi_L,\tilde{\delta}_{\pi_L})=V(\pi_L)$，因此 $\tilde{\delta}_{\pi_L}$ 是关于 π_L 的贝叶斯准则。不过，条件风险依赖于边界，即

$$R_j(\tilde{\delta}_{\pi_L})=qR_j(\delta^-_{\pi_L})+(1-q)R_j(\delta^+_{\pi_L}) \tag{2.3.12}$$

如果 q 按照下式选取，则可使条件风险 $R_0(\tilde{\delta}_{\pi_L})=R_1(\tilde{\delta}_{\pi_L})$最小，即

$$q=\frac{R_0(\delta^+_{\pi_L})-R_1(\delta^+_{\pi_L})}{R_0(\delta^+_{\pi_L})-R_1(\delta^+_{\pi_L})+R_1(\delta^-_{\pi_L})-R_0(\delta^-_{\pi_L})} \tag{2.3.13}$$

与命题(2.3.1)类似，按照式(2.3.13)选取 q 后的 $\tilde{\delta}_{\pi_L}$ 是极小化极大准则。

因为 V 是凹函数，所以在 π_L 处必然存在左导数 $V'(\pi_L^-)$ 和右导数 $V'(\pi_L^+)$。又因为 $V'(\pi_L^+)=[R_0(\delta_{\pi_L}^+)-R_1(\delta_{\pi_L}^+)]$ 和 $V'(\pi_L^-)=[R_0(\delta_{\pi_L}^-)-R_1(\delta_{\pi_L}^-)]$ 成立，所以可将式(2.3.12)改写为

$$q=\frac{V'(\pi_L^+)}{V'(\pi_L^+)-V'(\pi_L^-)} \tag{2.3.14}$$

注意，如果 $\{y\in\Gamma\mid(1-\pi_L)(C_{11}-C_{01})p_1(y)=\pi_L(C_{00}-C_{10})p_0(y)\}$ 在假设 H_0 和 H_1 下的概率均为 0，则判决准则为 δ_{π_L} 及其变化形式，无需采用 $\tilde{\delta}_{\pi_L}$。判决准则 $\tilde{\delta}_{\pi_L}$ 是随机判决准则的典型例子，与之相关的概念将在 2.4 节中继续讨论。

图 2.3.4 给出了判决准则 $\tilde{\delta}_{\pi_L}$ 的特性。直线 $r(\pi_0,\delta_{\pi_L}^-)$ 和 $r(\pi_0,\delta_{\pi_L}^+)$ 在 $\pi_0=\pi_L$ 处相交，斜率分别等于 $V'(\pi_0^-)$ 和 $V'(\pi_0^+)$。将概率 q 的值从 0 变到 1，便可得到包含在这两条直线间的任意直线。特别地，由式(2.3.14)确定的 q 所对应的直线是位于这两条直线之间的水平线。与图 2.3.2 的讨论类似，该判决准则显然具有极小化极大值特性。

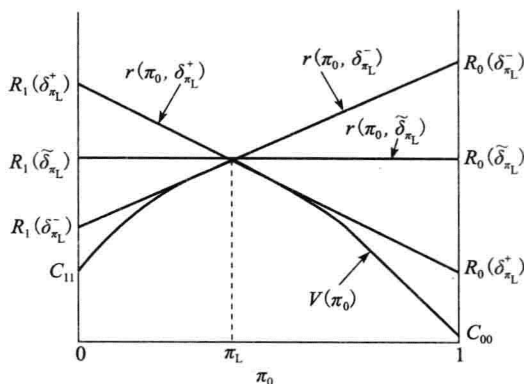


图 2.3.4　随机化判决准则

为了阐明极小化极大判决准则的设计思路，重新考虑 2.2 节中的两个例子。

例 2.3.1　高斯误差下的位置检验

考虑例 2.2.2 中含高斯误差且指定均匀代价的位置检验问题。函数 V 满足式(2.2.5)和式(2.2.31)，并由下式给出

$$V(\pi_0)=\pi_0\left(1-\Phi\left(\frac{\tau'-\mu_0}{\sigma}\right)\right)+(1-\pi_0)\Phi\left(\frac{\tau'-\mu_1}{\sigma}\right) \tag{2.3.15}$$

其中

$$\tau'=\frac{\sigma^2}{\mu_1-\mu_0}\log\left(\frac{\pi_0}{1-\pi_0}\right)+\frac{\mu_1+\mu_0}{2}$$

因为 $V(0)=C_{11}=0=C_{00}=V(1)$，所以最不利的先验概率 π_L 在区间(0，1)内。又因为式(2.3.15)是 π_0 的可微函数，所以不需要随机化处理，进而 π_L 可以通过 $R_0(\delta_{\pi_L})=R_1(\delta_{\pi_L})$ 得到。(事实上，因为 $L(Y)$ 是连续随机变量，对于任意 τ 都有 $P_0(L(Y)=\tau)=P_1(L(Y)=\tau)=0$ 成立，所以也不需要进行随机化处理。)先验概率 π_0 通过 τ' 影响判决风险 $R_0(\delta_{\pi_0})$ 和 $R_1(\delta_{\pi_0})$，因此通过求解下式可以得到均衡判决准则

$$1-\Phi\left(\frac{\tau'-\mu_0}{\sigma}\right)=\Phi\left(\frac{\tau'-\mu_1}{\sigma}\right) \tag{2.3.16}$$

结合图 2.3.5，上式存在唯一解，即

$$\tau_L'=\frac{\mu_0+\mu_1}{2} \tag{2.3.17}$$

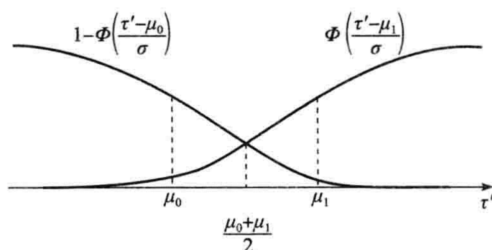


图 2.3.5　高斯误差和均匀代价下位置检验的条件风险

对应的极小化极大判决准则为

$$\delta_{\pi_L}(y)=\begin{cases}1, & y\geqslant(\mu_0+\mu_1)/2\\ 0, & y<(\mu_0+\mu_1)/2\end{cases} \tag{2.3.18}$$

从式(2.3.17)可知，最不利的先验概率是 $\pi_L=1/2$，对应的极小化极大风险为

$$V(1/2)=1-\Phi\left(\frac{\mu_1-\mu_0}{2\sigma}\right)$$

◀

例 2.3.2　二元信道

在第二个例子中，我们重新考虑指定均匀代价后例 2.2.1 所描述的二元信道问题。最小贝叶斯风险函数由下式给出(参见习题 1)

$$V(\pi_0)=\min\{(1-\pi_0)\lambda_1,\pi_0(1-\lambda_0)\}+\min\{(1-\pi_0)(1-\lambda_1),\pi_0\lambda_0\} \tag{2.3.19}$$

该函数可重写成

$$V(\pi_0)=\begin{cases}\pi_0, & 0\leqslant\pi_0\leqslant\underline{\pi}\\ \underline{\pi}+C\pi_0, & \underline{\pi}<\pi_0<\overline{\pi}\\ 1-\pi_0, & \overline{\pi}\leqslant\pi_0\leqslant 1\end{cases} \tag{2.3.20}$$

其中

$$\underline{\pi}=\min\left\{\frac{\lambda_1}{1-\lambda_0+\lambda_1},\frac{1-\lambda_1}{1-\lambda_1+\lambda_0}\right\},\quad \overline{\pi}=\max\left\{\frac{\lambda_1}{1-\lambda_0+\lambda_1},\frac{1-\lambda_1}{1-\lambda_1+\lambda_0}\right\}$$

并且 $C=(1-\overline{\pi}-\underline{\pi})/(\overline{\pi}-\underline{\pi})$。注意，斜率在 $\underline{\pi}$ 和 $\overline{\pi}$ 处发生改变，即 $V(\pi_0)$ 具有分段线性特性。又因为 $V(0)=V(1)=0$，所以在 $\underline{\pi}$ 和 $\overline{\pi}$ 之间斜率符号的变化将决定 π_L。特别地，如果 $C<0$，则 $\pi_L=\underline{\pi}$；如果 $C>0$，则 $\pi_L=\overline{\pi}$；如果 $C=0$，则 π_L 是先验概率 $[\underline{\pi},\overline{\pi}]$ 中的任意值。不论何种情况，其极小化极大风险都是 $V(\pi_L)=\max\{\underline{\pi},1-\overline{\pi}\}$。

无论是在 $\underline{\pi}$ 还是 $\overline{\pi}$，$V(\pi_0)$ 都不可微，所以有必要考虑随机化检验。为了方便说明，考虑图 2.3.6 中 $C>0$ 的情况。显然有 $\pi_L=\overline{\pi}$、$V'(\pi_L^+)=-1$ 和 $V'(\pi_L^-)=C$，进而可得到随机化常量 q 为

$$q=\frac{1}{1+C}=\frac{\overline{\pi}-\underline{\pi}}{1-2\underline{\pi}} \tag{2.3.21}$$

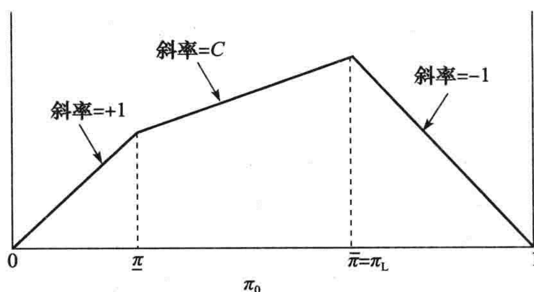


图 2.3.6　二元信道中的 $V(\pi_0)$

由例 2.2.1 易知：如果 $\pi_0>\overline{\pi}$，则贝叶斯准则是 $\delta_{\pi_0}(0)=\delta_{\pi_0}(1)=0$。因此 $\delta_{\pi_L}^+(0)=\delta_{\pi_L}^+(1)=0$ 和 Γ_1^+ 是空集。类似地，对于 $\underline{\pi}<\pi_0<\overline{\pi}$ 的情况，如果 $\lambda_1<1-\lambda_0$，则有 $\delta_{\pi_0}(y)=y$；如果 $\lambda_1>1-\lambda_0$，则有 $\delta_{\pi_0}(y)=1-y$。例如：如果 $\lambda_1<1-\lambda_0$，可以得到 $\delta_{\pi_L}^-(y)=y$，并且当 $y=0$ 时，极小化极大准则 $\tilde{\delta}_{\pi_L}$ 选择 0；当 $y=1$ 时，极小化极大准则 $\tilde{\delta}_{\pi_L}$ 以概率 q 选择 1[$y=0$ 时，以概率 $1-q$ 选择 0]。在这种情况下，极小化极大风险是

$$V(\overline{\pi}_L)=1-\overline{\pi}=\frac{\lambda_0}{1-\lambda_1+\lambda_0} \tag{2.3.22}$$

进而，如果 $\lambda_0=\lambda_1=\lambda$(若 $\lambda<1/2$，则 $\lambda_1<1-\lambda_0$)，有 $q=1$，且极小化极大代价为

$$V(\pi_L)=\lambda$$

就 $\pi_0\in[\lambda, 1-\lambda]$而言，该代价等于贝叶斯风险。换言之，只要 $\lambda\leqslant\pi_0\leqslant1-\lambda$，先验概率的不确定性不会影响判决性能。◀

综上所述，在先验概率未知的情况下，可采用极小化极大准则来设计最优判决器。该准则的解是关于最不利先验概率的贝叶斯检验，且通过在边界处进行随机化判决来获得极小化极大风险特性。此外，似然比检验通常具有最优判决特性。下一节将讨论二元假设检验问题的尼曼-皮尔逊假设检验。

2.4 尼曼-皮尔逊假设检验

针对 2.2 节描述的假设检验问题，贝叶斯准则通过最小化平均风险来设计最优判决器；类似地，极小化极大准则通过极小化极大条件风险来设计最优判决器。在很多实际问题中，对判决赋以正确代价通常较为困难。在这种情况下，常采用奈曼-皮尔逊判决准则作为替代的设计准则。

式(2.2.1)所描述的关于 H_0 和 H_1 的判决问题中存在两类误判：H_0 被错误接受或 H_1 被错误拒绝。第一类误判称为Ⅰ型误差或“虚警”；第二种误判称为Ⅱ型误差或“漏检”。“虚警”和“漏检”的提法来自于雷达领域，在雷达信号处理中，H_0 和 H_1 常用来代表不存在和存在目标，正确接受 H_1 通常称为检测。对于判决准则 δ，Ⅰ型误差的概率称为虚警概率（或虚警率），记为 $P_F(\delta)$。类似地，Ⅱ型误差称为漏检概率，用 $P_M(\delta)$来标记。但是，通常以检测概率来代替漏检概率，即 $P_D(\delta)=1-P_M(\delta)$，检测概率也称为判决 δ 的势。

显然，如果以牺牲某一个误差为代价，则可将上述两个误差中的任意一个设计得非常小。因此，设计 H_0和 H_1 的判决准则需要在这两种误差概率之间寻求平衡。事实上，贝叶斯准则和极小化极大准则本身就是种折衷处理方式。为了实现这种平衡，尼曼-皮尔逊准则首先固定虚警概率，然后通过最小化漏检概率来实现判决，即

$$\max_{\delta} P_D(\delta) \qquad \text{s.t.} \quad P_F(\delta)\leqslant\alpha \tag{2.4.1}$$

其中，α 是虚警概率约束，也称为检验水平或检验的显著性水平。因此，尼曼-皮尔逊准则设计的目标就是找到 H_0和 H_1 检验的显著水平为 α 的最大势判决。与贝叶斯准则和极小化极大准则不同，尼曼-皮尔逊准则假定两种假设的重要性存在差异。

为了给出式(2.4.1)所描述的尼曼-皮尔逊问题的通解，需要考虑与检验$\tilde{\delta}_{\pi_L}$类似的随机化检验，该检验的定义参见 2.3 节中极小化极大问题的解。从便于分析角度出发，可将随机化准则$\tilde{\delta}$定义为Γ 映射到区间[0，1]上的函数。换言之，对于 $y\in\Gamma$，$\tilde{\delta}(y)$是当观察结果为 $Y=y$ 时，接受 H_1 的条件概率。

例如，2.3 节介绍的随机化极小化极大准则$\tilde{\delta}_{\pi_L}$可写成

$$\tilde{\delta}_{\pi_L}(y)=\begin{cases}1, & L(y)>\tau_L\\ q, & L(y)=\tau_L\\ 0, & L(y)<\tau_L\end{cases} \tag{2.4.2}$$

式中，τ_L 是对应于最不利先验概率 π_L 的判决门限。根据这个定义，可以看出非随机化判决准则仅是随机化准则的特例。特别地，非随机化判决准则 δ 对应于随机化准则$\tilde{\delta}(y)=\delta(y)$。这两种方法的区别在于，$\delta$ 的值（0 或 1）是所接受假设的下标值，而$\tilde{\delta}$的值是接受

H_1 的概率。在$\tilde{\delta}$只取 0 或 1 时，这两种方法一致。

判决准则的虚警概率是当 H_0 为真的情况下接受 H_1 的概率。对于随机化判决准则$\tilde{\delta}$，其虚警概率根据下式计算

$$P_F(\tilde{\delta}) = E_0\{\tilde{\delta}(Y)\} = \int_\Gamma \tilde{\delta}(y)p_0(y)\mu(\mathrm{d}y) \tag{2.4.3}$$

式中，$E_0\{\cdot\}$表示假设 H_0 下的数学期望。式(2.4.3)之所以成立，是因为虚警概率与 H_0 假设下将观测 Y 判决为 H_1[即$\tilde{\delta}(Y)$]的平均概率相等。类似地，随机化选择的检测概率按照下式计算

$$P_D(\tilde{\delta}) = E_1\{\tilde{\delta}(Y)\} = \int_\Gamma \tilde{\delta}(y)p_1(y)\mu(\mathrm{d}y) \tag{2.4.4}$$

至此，尼曼-皮尔逊设计问题的通解可总结成以下结论。

命题 2.4.1　尼曼-皮尔逊引理

对于式(2.2.1)所描述的假设检验问题，给定 P_j 在 $j=0$ 和 $j=1$ 时的概率密度为 p_j 并假设 $\alpha>0$ 后，有如下结论成立：

1) 最优性。假设$\tilde{\delta}$是满足 $P_F(\tilde{\delta})\leqslant\alpha$ 的任意判决准则，$\tilde{\delta}'$为具有如下形式的任意判决准则

$$\tilde{\delta}'(y) = \begin{cases} 1, & p_1(y) > \eta p_0(y) \\ \gamma(y), & p_1(y) = \eta p_0(y) \\ 0, & p_1(y) < \eta p_0(y) \end{cases} \tag{2.4.5}$$

式中，$\eta\geqslant 0$，$\gamma(y)\in[0\quad 1]$满足 $P_F(\tilde{\delta}')=\alpha$，从而有 $P_D(\tilde{\delta}')\geqslant P_D(\tilde{\delta})$成立。即形如式(2.4.5)的任意判决均为虚警概率为 α 的尼曼-皮尔逊判决。

2) 存在性。对于所有的 $\alpha\in(0, 1)$，存在 $\gamma(y)=\gamma_0$ 的形如式(2.4.5)的判决准则$\tilde{\delta}_{NP}$，满足 $P_F(\tilde{\delta}_{NP})=\alpha$。

3) 唯一性。假设$\tilde{\delta}''$是任一虚警概率为 α 的尼曼-皮尔逊判决准则，则$\tilde{\delta}''$必然具有式(2.4.5)的形式，除非 H_0 和 H_1 假设在 Γ 的某个子集上的概率均为 0。

证明　1) 假设$\tilde{\delta}$和$\tilde{\delta}'$的定义如前所述。在给定$\tilde{\delta}'$后，对于 $y\in\Gamma$，$[\tilde{\delta}'(y)-\tilde{\delta}(y)]\cdot[p_1(y)-\eta p_0(y)]\geqslant 0$ 总是成立的。因此

$$\int_\Gamma [\tilde{\delta}'(y) - \tilde{\delta}(y)][p_1(y) - \eta p_0(y)]\mu(\mathrm{d}y) \geqslant 0 \tag{2.4.6}$$

将式(2.4.6)中各项展开并重排后，有

$$\int_\Gamma \tilde{\delta}' p_1 \mathrm{d}\mu - \int_\Gamma \tilde{\delta} p_1 \mathrm{d}\mu \geqslant \eta\left[\int_\Gamma \tilde{\delta}' p_0 \mathrm{d}\mu - \int_\Gamma \tilde{\delta} p_0 \mathrm{d}\mu\right] \tag{2.4.7}$$

利用式(2.4.3)和式(2.4.4)，可将式(2.4.7)改写成

$$P_D(\tilde{\delta}') - P_D(\tilde{\delta}) \geqslant \eta[P_F(\tilde{\delta}') - P_F(\tilde{\delta})] = \eta[\alpha - P_F(\tilde{\delta})] \geqslant 0 \tag{2.4.8}$$

在推导过程中利用了 $P_F(\tilde{\delta})\leqslant\alpha$。至此，完成了 $P_D(\tilde{\delta}')\geqslant P_D(\tilde{\delta})$的证明。

2) 如果假设 η_0是使下式成立的最小常数

$$P_0(p_1(Y) > \eta_0 p_0(Y)) \leqslant \alpha \tag{2.4.9}$$

那么在 $P_0(p_1(Y)>\eta_0 p_0(Y))<\alpha$ 时，可选择

$$\gamma_0 = \frac{\alpha - P_0(p_1(Y) > \eta_0 p_0(Y))}{P_0(p_1(Y) = \eta_0 p_0(Y))} \tag{2.4.10}$$

其他情况可任意选取 γ_0，具体情况如图 2.4.1 所示。进一步假设 $\tilde{\delta}_{NP}$ 是 $\eta=\eta_0$ 和 $\gamma(y)=\gamma_0$ 时由式(2.4.5)所描述的判决准则。根据定义，有

$$\begin{aligned}P_F(\tilde{\delta}_{NP}) &= E_0\{\tilde{\delta}_{NP}(Y)\}\\ &= P_0(p_1(Y) > \eta_0 p_0(Y))\\ &\quad + \gamma_0 P_0(p_1(Y)\\ &= \eta_0 p_0(Y)) = \alpha \end{aligned} \quad (2.4.11)$$

显然，存在形如式(2.4.5)所描述的判决准则，使得 $\gamma(y)$ 为常数，且虚警概率为 α。

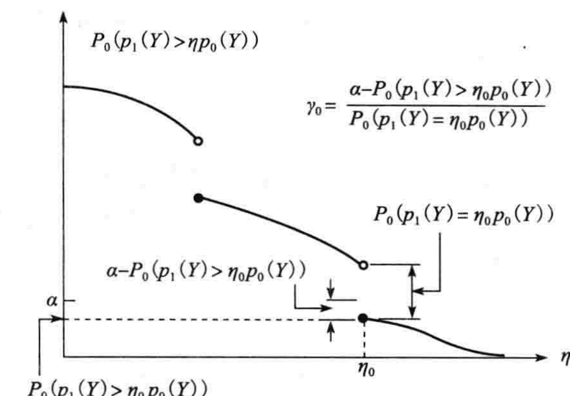


图 2.4.1 尼曼-皮尔逊假设检验的判决门限和随机化参数

3) 假设 $\tilde{\delta}'$ 为形如式(2.4.5)的虚警概率为 α 的尼曼-皮尔逊判决准则，$\tilde{\delta}''$ 为任一种另外的虚警概率为 α 的尼曼-皮尔逊判决准则。根据定义，则 $P_D(\tilde{\delta}'')$ 必然等于 $P_D(\tilde{\delta}')$，进而结合式(2.4.8)可得 $0\geqslant\alpha-P_F(\tilde{\delta}'')\geqslant 0$，即 $P_F(\tilde{\delta}'')=\alpha$。因此，$\tilde{\delta}''$ 的大小必为 α。利用等式 $P_D(\tilde{\delta}'')=P_D(\tilde{\delta}')$ 和 $P_F(\tilde{\delta}'')=P_F(\tilde{\delta}')$，并结合式(2.4.6)～式(2.4.8)的推导过程，可以得到

$$\int_\Gamma [\tilde{\delta}'(y)-\tilde{\delta}''(y)][p_1(y)-\eta p_0(y)]\mu(\mathrm{d}y)=0 \quad (2.4.12)$$

由于被积函数非负，因此除了概率为 0 的集合之外，式(2.4.12)也应该等于 0。因此，$\tilde{\delta}'$ 和 $\tilde{\delta}''$ 仅在集合 $\{y\in\Gamma \mid p_1(y)=\eta p_0(y)\}$ 中存在差异。换言之，$\tilde{\delta}''$ 也具有式(2.4.5)的形式，与 $\tilde{\delta}'$ 的区别仅在于函数 $\gamma(y)$ 的形式可能不同。 □

上述结果再次证明似然比检验具有最优性。对某个给定的假设检验对，尼曼-皮尔逊检验与贝叶斯和极小化极大检验的差异仅在于判决门限和随机化参数。(注意，当 $\alpha=0$ 时，尼曼-皮尔逊检验可由非随机化检验直接给出，其关键域为 $\Gamma_1=\{y\in\Gamma \mid p_0(y)=0\}$)。下面结合两个例子来说明尼曼-皮尔逊检验的设计方法。

例 2.4.1 高斯误差下的位置检验

再次考虑例 2.2.2 所描述的高斯误差下的位置检验问题，有

$$P_0(p_1(Y)>\eta_0 p_0(Y))=P_0(L(Y)>\eta)=P_0(Y>\eta')=1-\Phi\left(\frac{\eta'-\mu_0}{\sigma}\right) \quad (2.4.13)$$

式中，$\eta'=\sigma^2\log(\eta)/(\mu_1-\mu_0)+(\mu_1+\mu_0)/2$。图 2.4.2 给出了虚警概率与判决门限的变化关系。

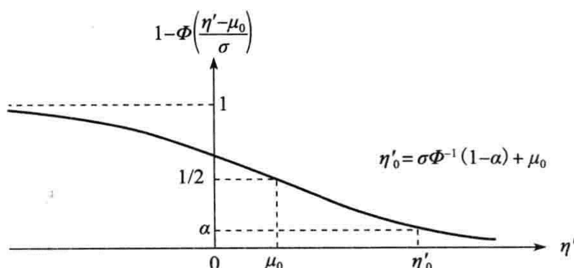


图 2.4.2 高斯误差下位置的尼曼-皮尔逊检验的判决门限

给定判决门限后，虚警概率按照下式计算

$$\eta'_0=\sigma\Phi^{-1}(1-\alpha)+\mu_0 \quad (2.4.14)$$

式中，Φ^{-1} 是 Φ 的逆函数。又因为 $P(Y=\eta'_0)=0$，这意味着可进行任意随机化处理，例如选择 $\gamma_0=1$。此时，虚警概率为 α 的尼曼-皮尔逊检验可由下式给出

$$\widetilde{\delta}_{\mathrm{NP}}(y)=\begin{cases}1, & y\geqslant \eta_0'\\ 0, & y<\eta_0'\end{cases} \tag{2.4.15}$$

式中，$\widetilde{\delta}_{\mathrm{NP}}$具有式(2.4.14)的形式。

根据定义，$\widetilde{\delta}_{\mathrm{NP}}$的检测概率为

$$\begin{aligned}P_{\mathrm{D}}(\widetilde{\delta}_{\mathrm{NP}})&=E_1\{\widetilde{\delta}_{\mathrm{NP}}(Y)\}=P_1(Y\geqslant\eta_0')=1-\Phi\left(\frac{\eta_0'-\mu_1}{\sigma}\right)\\&=1-\Phi(\Phi^{-1}(1-\alpha)-d)\end{aligned} \tag{2.4.16}$$

式中，$d=(\mu_1-\mu_0)/\sigma$ 是例 2.2.2 中定义的信噪比。对于固定的 α，由式(2.4.16)得到检测概率是信噪比 d 的函数。这种关系有时也称作检验的势函数，如图 2.4.3 所示。给定 d 后，式(2.4.16)给出了检测概率与虚警概率的关系。借用雷达技术中的术语，称检测概率与虚警概率的关系为接收机工作特性(Receiver Operating Characteristics，ROC)，如图 2.4.4 所示。图 2.4.2～图 2.4.4 可用于分析高斯噪声中的信号检测问题。后续章节将深入讨论与之相关的内容。◀

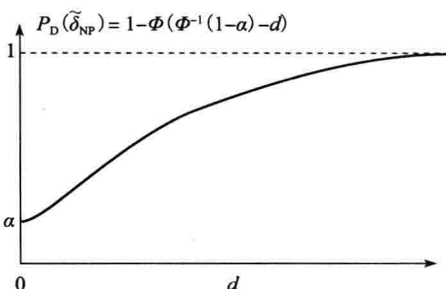


图 2.4.3　高斯误差($d=(\mu_1-\mu_0)/\sigma$)下位置的尼曼-皮尔逊检验的检测概率

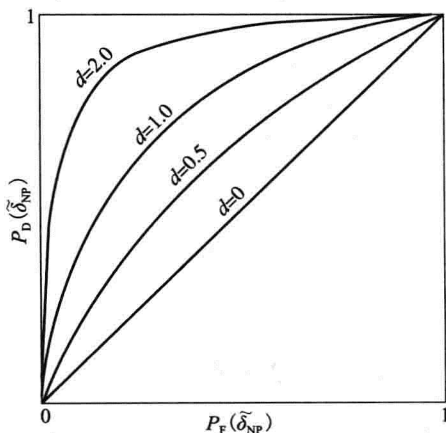


图 2.4.4　高斯误差($d=(\mu_1-\mu_0)/\sigma$)下位置的尼曼-皮尔逊检验的 ROC 曲线

例 2.4.2　二元信道

为了说明随机化在尼曼-皮尔逊问题中的重要性，再次讨论例 2.2.1 中的二元信道问题。尽管式(2.2.20)给出了此类问题的似然比，但就尼曼-皮尔逊检测而言，在求解虚警概率 α 对应的判决门限时，需要考虑 $P_0(L(Y)>\eta)$。为了简化分析，假设 $\lambda_0+\lambda_1<1$，进而有 $\lambda_1/(1-\lambda_0)<(1-\lambda_1)/\lambda_0$ 成立，并且

$$P_0(L(Y)>\eta)=\begin{cases}1, & \eta<\dfrac{\lambda_1}{1-\lambda_0}\\ \lambda_0, & \dfrac{\lambda_1}{1-\lambda_0}\leqslant\eta<\dfrac{1-\lambda_1}{\lambda_0}\\ 0, & \eta\geqslant\dfrac{1-\lambda_1}{\lambda_0}\end{cases} \tag{2.4.17}$$

该函数如图 2.4.5 所示。结合图 2.4.5 易知，虚警概率为 α 的尼曼-皮尔逊判决器的判决门限为

$$\eta_0=\begin{cases}\dfrac{1-\lambda_1}{\lambda_0}, & 0\leqslant\alpha<\lambda_0\\ \dfrac{\lambda_1}{1-\lambda_0}, & \lambda_0\leqslant\alpha<1\\ 0, & \alpha=1\end{cases} \tag{2.4.18}$$

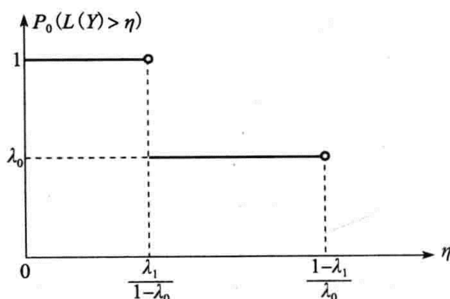


图 2.4.5　二元信道的阈值和随机化选择

而且，还可以看出相应的随机化常量为

$$\gamma_0=\begin{cases}\dfrac{\alpha}{\lambda_0}, & 0\leqslant\alpha<\lambda_0\\ \dfrac{\alpha-\lambda_0}{1-\lambda_0}, & \lambda_0\leqslant\alpha<1\\ \text{任意值}, & \alpha=1\end{cases}\tag{2.4.19}$$

结合式(2.2.20)、式(2.2.18)和式(2.2.19)可知，在 $0\leqslant\alpha<\lambda_0$ 时，尼曼-皮尔逊假设检验 $\tilde{\delta}_{NP}(y)$ 为

$$\tilde{\delta}_{NP}(y)=\begin{cases}\dfrac{\alpha}{\lambda_0}, & y=1\\ 0, & y=0\end{cases}\tag{2.4.20}$$

当 $\lambda_0\leqslant\alpha\leqslant1$ 时，$\tilde{\delta}_{NP}(y)$ 为

$$\tilde{\delta}_{NP}(y)=\begin{cases}1, & y=1\\ \dfrac{\alpha-\lambda_0}{1-\lambda_0}, & y=0\end{cases}\tag{2.4.21}$$

根据 $P_D(\tilde{\delta}_{NP})=P_1(L(Y)>\eta_0)+\gamma_0P_1(L(Y)=\eta_0)$ 可计算尼曼-皮尔逊检验的检测概率，具体为

$$P_D(\tilde{\delta}_{NP})=\begin{cases}\alpha\dfrac{1-\lambda_1}{\lambda_0}, & 0\leqslant\alpha<\lambda_0\\ (1-\lambda_1)+\lambda_1\dfrac{\alpha-\lambda_0}{1-\lambda_0}, & \lambda_0\leqslant\alpha\leqslant1\end{cases}\tag{2.4.22}$$

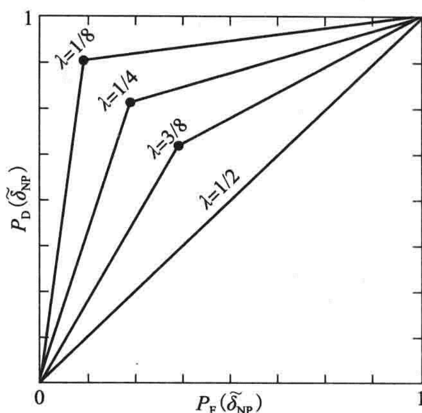


图 2.4.6　对称二元信道的 ROC

值得注意的是：ROC(P_D 关于 P_F)分段线性，在 $P_F=\lambda_0$ 处斜率发生变化。图 2.4.6 给出了对称信道($\lambda_0=\lambda_1=\lambda\leqslant1/2$)的 ROC 曲线。 ◀

2.5 复合假设检验

前面几节讨论的假设检验问题也称作简单假设检验问题，这是因为式(2.2.1)所描述的两个假设均只对应观测数据的单个分布。但在许多假设检验问题中，每个假设中可以出现多种可能的分布。这种类型的假设称为复合假设。例如，在雷达检测问题中，回波信号存在未知的参数，如到达时间(与位置有关)和多普勒频移(与速度有关)。因此，“目标存在”的假设将是一个复合问题。本节将简要讨论复合假设检验的判决准则设计方法。

为了不失一般性，对复合假设检验问题进行建模，需要考虑在 Γ 上以参数 θ 表示的概率密度族，其中 $\theta\in\Lambda$。即有一族概率分布$\{P_\theta;\ \theta\in\Lambda\}$，其中，$P_\theta$ 是 θ 为真实参数时观测数据的概率。Λ 中参数的值代表了参数所有可能取值的集合。式(2.2.1)所描述的简单假设对应 $\Lambda=\{0,\ 1\}$的特殊情况。更一般地，参数空间可由两个不相交的集合 Λ_0 和 Λ_1 构成，它们分别对应两种假设下参数的值域。

在复合假设检验问题的贝叶斯判决形式中，假设参数是在集合 Λ 上取值的随机变量 Θ。此时，P_θ 表示给定 $\Theta=\theta$ 时 Y 的条件分布，期望对 Θ 进行二元判决。为了便于分析，

仅考虑 Θ 的非随机化二元判决准则。在贝叶斯意义上设计最优判决准则，首先需要指定代价函数 $C[i,\theta]$，其中，$C[i,\theta]$ 表示当 $Y\sim P_\theta$ 时($i\in\{0,1\}$，$\theta\in\Lambda$)接受假设 i 的代价。例如，在雷达检测问题中，参数 θ 可能与目标位置及速度有关。相比距离较远的慢速目标而言，丢失邻近高速目标的代价可能更大。为了分析简便，限定 C 非负且有界。

与简单情况类似，判决准则 δ 的条件风险定义为

$$R_\theta(\delta)=E_\theta\{C[\delta(Y),\theta]\},\quad \theta\in\Lambda \tag{2.5.1}$$

式中，E_θ 表示 $Y\sim P_\theta$ 的期望。类似地，平均或贝叶斯风险可以定义为

$$r(\delta)=E\{R_\Theta(\delta)\} \tag{2.5.2}$$

最小化 $r(\delta)$ 可得到贝叶斯判决准则。

考虑 $E_\theta\{C[\delta(Y),\theta]\}\triangleq E\{C[\delta(Y),\Theta]\,|\,\Theta=\theta\}$ 后，$r(\delta)$ 可写成

$$r(\delta)=E\{E\{C[\delta(Y),\Theta]\,|\,\Theta\}\}=E\{C[\delta(Y),\Theta]\} \tag{2.5.3}$$

其中，第二个等式由 $E\{X\}=E\{E\{X|Y\}\}$ 得到。因此，$r(\delta)$ 是判决准则 δ 关于 Θ 和 Y 的平均代价。再次利用 $E\{X\}=E\{E\{X|Y\}\}$ 后，可以得到

$$r(\delta)=E\{E\{C[\delta(Y),\Theta]\,|\,Y\}\} \tag{2.5.4}$$

在式(2.5.4)中需要注意的是：如果对于每个 $y\in\Gamma$，判决 $\delta(y)$ 都能够最小化下式所描述的后验代价，那么判决 δ 能最小化 $r(\delta)$

$$E\{C[\delta(Y),\Theta]\,|\,Y=y\} \tag{2.5.5}$$

又因为 $\delta(y)$ 只能取 0 或 1，所以可得到贝叶斯判决准则如下

$$\delta_{\mathrm{B}}(y)=\begin{cases}1, & E\{C[1,\Theta]\,|\,Y=y\}<E\{C[0,\Theta]\,|\,Y=y\}\\ 0\text{ 或 }1, & E\{C[1,\Theta]\,|\,Y=y\}=E\{C[0,\Theta]\,|\,Y=y\}\\ 0, & E\{C[1,\Theta]\,|\,Y=y\}>E\{C[0,\Theta]\,|\,Y=y\}\end{cases} \tag{2.5.6}$$

式(2.5.6)的解释很直观：在给定观测数据后，δ_{B} 选择了平均代价最小的假设。在 $\Lambda=\{0,1\}$ 时，式(2.5.6)简化成 2.2 节讨论过的简单假设检验的贝叶斯规则，同样可理解为最小化后验代价。

在许多问题中，参数空间可以分解为两个具有均匀代价的不相交集合 Λ_0 和 Λ_1，这两个集合分别代表假设 H_0 和 H_1，即

$$C[i,\theta]=C_{ij},\quad \theta\in\Lambda_j \tag{2.5.7}$$

显然，假设 $C_{11}<C_{01}$ 后，式(2.5.6)可以简化为

$$\delta_{\mathrm{B}}(y)=\begin{cases}1, & \dfrac{P(\Theta\in\Lambda_1\,|\,Y=y)}{P(\Theta\in\Lambda_0\,|\,Y=y)}>\dfrac{C_{10}-C_{00}}{C_{01}-C_{11}}\\[2ex] 0\text{ 或 }1, & \dfrac{P(\Theta\in\Lambda_1\,|\,Y=y)}{P(\Theta\in\Lambda_0\,|\,Y=y)}=\dfrac{C_{10}-C_{00}}{C_{01}-C_{11}}\\[2ex] 0, & \dfrac{P(\Theta\in\Lambda_1\,|\,Y=y)}{P(\Theta\in\Lambda_0\,|\,Y=y)}<\dfrac{C_{10}-C_{00}}{C_{01}-C_{11}}\end{cases} \tag{2.5.8}$$

式中，$P(\Theta\in\Lambda_j\,|\,Y=y)$ 表示 $Y=y$ 时，Θ 位于 Λ_j 的条件概率。进一步假设 Y 的条件概率密度为 $p(y|\Theta\in\Lambda_j)$，$j=0,1$，利用贝叶斯公式可以得到

$$P(\Theta\in\Lambda_j\,|\,Y=y)=\frac{p(y\,|\,\Theta\in\Lambda_j)P(\Theta\in\Lambda_j)}{p(y)} \tag{2.5.9}$$

式中，$p(y)=\sum_{j=0}^{1}p(y\,|\,\Theta\in\Lambda_j)P(\Theta\in\Lambda_j)$。进而可将式(2.5.8)简化为

$$\delta_B(y)=\begin{cases}1, & L(y)<\dfrac{\pi_0(C_{10}-C_{00})}{\pi_1(C_{01}-C_{11})}\\ 0\text{ 或 }1, & L(y)=\dfrac{\pi_0(C_{10}-C_{00})}{\pi_1(C_{01}-C_{11})}\\ 0, & L(y)>\dfrac{\pi_0(C_{10}-C_{00})}{\pi_1(C_{01}-C_{11})}\end{cases}\tag{2.5.10}$$

式中，$\pi_j=P(\Theta\in\Lambda_j)$，$L(y)=p(y|\Theta\in\Lambda_1)/p(y|\Theta\in\Lambda_0)$。从式(2.5.10)可以看出，在$p_j(y)=p(y|\Theta\in\Lambda_j)$时，该问题与简单贝叶斯假设检验问题等价。

对于所有的$\theta\in\Lambda$，假设Θ具有密度$w(\theta)$，P_θ具有密度p_θ，利用第1章中的符号说明，密度$p(y|\Theta\in\Lambda_j)$由下式给出

$$p(y|\Theta\in\Lambda_j)=\int_\Lambda p_\theta(y)w_j(\theta)\mu(\mathrm{d}\theta)\tag{2.5.11}$$

式中，$w_j(\theta)$是$\Theta\in\Lambda_j$时Θ的条件密度，即

$$w_j(\theta)=\begin{cases}0, & \theta\notin\Lambda_j\\ w(\theta)/\pi_j, & \theta\in\Lambda_j\end{cases}\tag{2.5.12}$$

式中，$\pi_j=\int_{\Lambda_j}w(\theta)\mu(\mathrm{d}\theta)$。

注意，该问题中的假设检验对也可以由式(2.5.11)中的观测密度定义，它只取决于条件密度w_j。因此，对于w_j已知、而π_j未知的复合假设检验问题，也可以采用极小化极大准则和尼曼-皮尔逊判决准则。当然，除了在参数模型的物理解释上存在差异外，它们与2.2节～2.4节中所描述的简单假设检验问题在本质上是一致的。下面通过例子来说明如何设计复合假设检验的判决准则。

例2.5.1 平面内点分布半径检验

假设$\Gamma=\mathbf{R}^2(Y=(Y_1, Y_2)^{\mathrm{T}})$，有如下假设

$$\begin{aligned}H_0:&\begin{aligned}Y_1&=\varepsilon_1\\ Y_2&=\varepsilon_2\end{aligned}\\ H_1:&\begin{aligned}Y_1&=A\cos\Psi+\varepsilon_1\\ Y_2&=A\sin\Psi+\varepsilon_2\end{aligned}\end{aligned}\tag{2.5.13}$$

式中，A是正常数，Ψ是在$[0, 2\pi]$上均匀分布的随机变量，ε_1和ε_2为服从$\mathcal{N}(0, \sigma^2)$分布的随机变量，ε_1、ε_2和Ψ相互独立。这里的观测数据可视为平面内某点坐标的带噪声测量值。该点可能位于原点，也可能在一个以A为半径的圆内均匀分布(第3章将会讨论有关该模型的一些应用)。

针对上述检测问题，参数可以取为$\Theta=(\Theta_1, \Theta_2)$，其中$\Theta_1\in\{0, A\}$，$\Theta_2\in[0, 2\pi]$。因此，参数集$\Lambda$为$\{0, A\}\times[0, 2\pi]$且$\Lambda_0=\{\theta\in\Lambda|\theta_1=0\}$，$\Lambda_1=\{\theta\in\Lambda|\theta_1=A\}$。当$\Theta=\theta$时，$Y$的密度函数为两个独立的$\mathcal{N}(0, \sigma^2)$分布分别进行均值平移$\theta_1\cos\theta_2$和$\theta_1\sin\theta_2$后的联合分布，即

$$p_\theta(y)=\frac{1}{2\pi\sigma^2}\exp\{-q(y,\theta)/2\sigma^2\},\quad y\in\mathbf{R}^2$$

$$q(y,\theta)\triangleq[(y-\theta_1\cos\theta_2)^2+(y-\theta_1\sin\theta_2)^2]\tag{2.5.14}$$

进而可以直接得到

$$p(y \mid \Theta \in \Lambda_0) = p_\theta(y) \mid_{\theta_1=0} = \frac{1}{2\pi\sigma^2}\exp\{-(y_1^2 + y_2^2)/2\sigma^2\} \tag{2.5.15}$$

和

$$\begin{aligned} p(y \mid \Theta \in \Lambda_1) &= \frac{1}{2\pi}\int_0^{2\pi} p_\theta(y) \mid_{\theta_1=A} \mathrm{d}\theta_2 \\ &= \frac{1}{4\pi^2\sigma^2}\int_0^{2\pi} \exp\{-q(y,\theta) \mid_{\theta_1=A}/2\sigma^2\}\mathrm{d}\theta_2 \end{aligned} \tag{2.5.16}$$

结合式(2.5.15)和式(2.5.16)可以得到似然比如下

$$L(y) = \frac{p(y \mid \Theta \in \Lambda_1)}{p(y \mid \Theta \in \Lambda_0)} = \frac{\mathrm{e}^{-A^2/2\sigma^2}}{2\pi}\int_0^{2\pi} \exp\left\{\frac{A}{\sigma^2}(y_1\cos\theta_2 + y_2\sin\theta_2)\right\}\mathrm{d}\theta_2 \tag{2.5.17}$$

上式利用了 $\sin^2\theta_2+\cos^2\theta_2=1$。为了简化式(2.5.17)，引入变量 $r=(y_1^2+y_2^2)^{1/2}$ 和变量 $\phi=\arctan(y_2/y_1)$ 后，有 $y_1=r\cos\phi$，$y_2=r\sin\phi$。利用三角公式 $\cos\phi\cos\theta_2+\sin\phi\sin\theta_2=\cos(\theta_2-\phi)$，可以得到

$$L(y) = \frac{\mathrm{e}^{-A^2/2\sigma^2}}{2\pi}\int_0^{2\pi} \exp\left\{\frac{Ar}{\sigma^2}\cos(\theta_2-\phi)\right\}\mathrm{d}\theta_2 = \mathrm{e}^{-A^2/2\sigma^2} I_0\left(\frac{Ar}{\sigma^2}\right) \tag{2.5.18}$$

式中，I_0 是第一类零阶修正贝塞尔函数，它对应式(2.5.18)中的积分项。

函数 $I_0(x)$ 单调递增。按照式(2.5.18)计算 $L(y)$ 并与门限 τ 进行比较等价于将 r 与 $\tau'=\sigma^2 I_0^{-1}(\tau\mathrm{e}^{A^2/2\sigma^2})/A$ 进行比较。式(2.5.13)所描述的检验问题的贝叶斯、极小化极大以及尼曼一皮尔逊判决准则都具有如下形式

$$\tilde{\delta}_0(y) = \begin{cases} 1, & r > \tau' \\ \gamma, & r = \tau' \\ 0, & r < \tau' \end{cases} \tag{2.5.19}$$

因此，可通过将 r 与门限进行比较来实现对式(2.5.13)的最优判决，其中 r 为点(y_1, y_2)到原点的距离。该判决域如图 2.5.1 所示(这里随机化方式不重要)。第 3 章将会进一步讨论该最优检验的其他方面及相关模型。◀

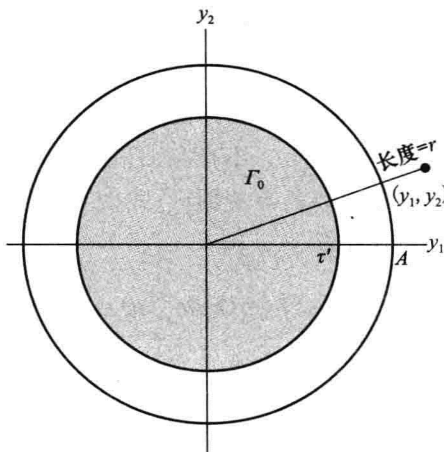


图 2.5.1　例 2.5.1 的判决域($\Gamma_1=\Gamma_0^c$)

对于复合假设检验问题，在缺乏参数的先验分布(或条件先验)时，要给出严格满足最优性定义的假设检验通常是很困难的。此时，需要对 2.4 节中尼曼-皮尔逊准则进行推广。与上节类似，假设参数空间可分为两个不相交的集合 Λ_0 和 Λ_1。对于任意随机化判决准则 $\tilde{\delta}$，可定义其虚警概率和检测概率为

$$\begin{aligned} P_{\mathrm{F}}(\tilde{\delta};\theta) &= E_\theta\{\tilde{\delta}(Y)\}, \quad \theta \in \Lambda_0 \\ P_{\mathrm{D}}(\tilde{\delta};\theta) &= E_\theta\{\tilde{\delta}(Y)\}, \quad \theta \in \Lambda_1 \end{aligned} \tag{2.5.20}$$

与尼曼-皮尔逊表达式类似，要求虚警概率不超过某个给定值 α。显然，满足 $P_{\mathrm{F}}(\tilde{\delta};\theta) \leqslant \alpha$，$\theta \in \Lambda_0$ 约束，并对每个 $\theta \in \Lambda_1$ 都能最大化 $P_{\mathrm{D}}(\tilde{\delta};\theta)$ 的判决准则将是最优的。具有上述特性的判决准则也称为虚警概率为 α 的一致最大势检验(Uniformly Most Powerful，UMP)。

尽管非常希望得到 UMP 检验，但它仅在某些特殊情况下才存在。为了说明这一问题，考虑单一零假设 H_0 情况，即 Λ_0 仅包含单个元素 θ_0。假设对每个 $\theta\in\Lambda$，P_θ 具有密度 p_θ，则 H_0 和假设 $Y\sim P_\theta$ 的虚警概率 α 的最大势检验的判决域为 $\Gamma_\theta=\{y\in\Gamma|p_\theta(y)>\tau p_{\theta_0}(y)\}$，其中 τ(也有可能是随机化)根据虚警概率 α 计算。另外，根据尼曼-皮尔逊引理，该检验具有唯一性，因此，其他任何虚警概率为 α 的检验都不会比它更有效。例如，对于 Λ_1 中的任意两个元素 θ' 和 θ''，就判决 H_0 和 $Y\sim P_{\theta'}$ 而言，基于 $\Gamma_{\theta'}$ 的检验比基于 $\Gamma_{\theta''}$ 的检验更有效(反之亦然)，除非两个关键域完全相同。因此，可以得出结论，当且仅当对所有的 $\theta\in\Lambda_1$，Γ_θ 都相同时，才存在 H_0 与复合假设 $H_1:Y\sim P_\theta$，$\theta\in\Lambda_1$ 的 UMP 检验。下面的例子将进一步说明这一问题。

例 2.5.2 位置的 UMP 检验

考虑分布$\{P_\theta;\ \theta\in\Lambda\}$的参数集，其中 P_θ 服从$\mathcal{N}(\theta,\sigma^2)$分布，$\Lambda$ 为 $\mathbf{R}$ 的一个子集。假设对为

$$\begin{aligned}H_0&:\theta=\mu_0\\ H_1&:\theta>\mu_0\end{aligned}\tag{2.5.21}$$

式中，μ_0 是固定常数。该问题是典型的单一零假设问题，即 $\Lambda_0=\{\mu_0\}\cup\Lambda_1=\{\mu_0,\ \infty\}$。从例 2.4.1 可知，对于每个 $\theta\in\Lambda$，H_0 与 $Y\sim\mathcal{N}(\theta,\sigma^2)$的虚警概率为 α 的最大势检验具有关键域 $\Gamma_\theta=\{y\in\Gamma|\,y>\sigma\Phi^{-1}(1-\alpha)+\mu_0\}$。由于该域不取决于 θ，因此，它给出了(2.5.21)的 UMP 检验，将其记作 $\tilde{\delta}_1$。注意，式(2.4.16)意味着

$$P_D(\tilde{\delta}_1;\theta)=1-\Phi\left(\Phi^{-1}(1-\alpha)-\frac{\theta-\mu_0}{\sigma}\right)\tag{2.5.22}$$

另外，对于相同的分布族，考虑如下假设对

$$\begin{aligned}H_0&:\theta=\mu_0\\ H_1&:\theta\neq\mu_0\end{aligned}\tag{2.5.23}$$

现在有复合假设 $\Lambda_1=(-\infty,\ \mu_0)\cup(\mu_0,\ \infty)$。对于 $\theta>\mu_0$，最大势的关键域同前。但是，对于 $\theta<\mu_0$，可以直接看出虚警概率为 α 的最大势检验的关键域为

$$\Gamma_\theta=\{y\in\Gamma|\,y<\sigma\Phi^{-1}(\alpha)+\mu_0\}\tag{2.5.24}$$

虽然该域与 θ 无关，但在 $\theta>\mu_0$ 时，它与 Γ_θ 截然不同。因此，对式(2.5.23)所描述的假设检验问题而言，不存在 UMP 检验。

如果将具有关键域[式(2.5.24)所示]的判决记作 $\tilde{\delta}_2$，则

$$P_D(\tilde{\delta}_2;\theta)=\Phi\left(\Phi^{-1}(\alpha)-\frac{\theta-\mu_0}{\sigma}\right)\tag{2.5.25}$$

图 2.5.2 给出了该量值以及式(2.5.22)中的 $P_D(\tilde{\delta}_1;\ \theta)$ 相对于 θ 的曲线图。注意，当 θ 在最优区域外时，两个假设的性能都不是很好。[对式(2.5.23)更合理的一个检验是将 $|y-\mu_0|$ 和门限进行比较，而不是 $\tilde{\delta}_1$ 或 $\tilde{\delta}_2$，但该检验不具有 UMP 特性。] ◀

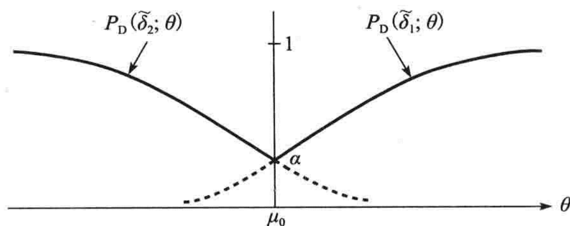


图 2.5.2 高斯误差下位置检验的检测概率
($\theta=\mu_0$、$\theta>\mu_0$ 和 $\theta=\mu_0$、$\theta<\mu_0$)

例 2.5.2 说明了在大多数情况下，UMP 准则都过于严格。采用其他约束后，可以消除一些不合理的检验。例如*无偏性约束*，它要求对于所有的 $\theta\in\Lambda_1$，$P_D(\tilde{\delta};\theta)\geqslant\alpha$，并且 $P_F(\tilde{\delta};\theta)\leqslant\alpha$。注意，该约束能同时消除式(2.5.23)中的检验 $\tilde{\delta}_1$ 和 $\tilde{\delta}_2$。与之相关的讨论可以参考 Lehmann(1986)的书。

在实际应用中，参数集 Λ 通常具有 $[\theta_0,\infty)$ 形式，其中 $\Lambda_0=\{\theta_0\}$，$\Lambda_1=(\theta_0,\infty)$；因此，有假设对

$$\begin{aligned} H_0&:\theta=\theta_0 \\ H_1&:\theta>\theta_0 \end{aligned} \tag{2.5.26}$$

上述判决可与 $\theta_0=0$、信号幅度为 θ 的信号检测问题对应。此时，在 H_1 假设下，我们更关心 θ 接近 θ_0 的情况，即弱信号判决问题。

对于判决准则 $\tilde{\delta}$，在正则性约束下，可以将 $P_D(\tilde{\delta};\theta)$ 展开成 θ_0 的泰勒级数，即

$$P_D(\tilde{\delta};\theta)=P_D(\tilde{\delta};\theta_0)+(\theta-\theta_0)P_D'(\tilde{\delta};\theta_0)+O((\theta-\theta_0)^2) \tag{2.5.27}$$

式中，$P_D'(\tilde{\delta};\theta_0)=\partial P_D(\tilde{\delta};\theta)/\partial\theta$。利用 $P_D'(\tilde{\delta};\theta_0)=P_F(\tilde{\delta})$，对虚警概率为 α 的判决准则 $P_D(\tilde{\delta};\theta)$ 而言，在 θ_0 的邻域内有下面的表达式近似成立

$$P_D(\tilde{\delta};\theta)\approx\alpha+(\theta-\theta_0)P_D'(\tilde{\delta};\theta_0) \tag{2.5.28}$$

因此，对于 θ_0 附近的 θ，可以通过最大化 $P_D'(\tilde{\delta};\theta_0)$ 来得到近似最大检测概率的判决。能够使 $P_D'(\tilde{\delta};\theta_0)$ 最大化，并具有虚警约束 $P_F(\tilde{\delta})\leqslant\alpha$ 的检验称为虚警概率为 α 的*局部最大势检验*(Locally Most Powerful，LMP)，或称为*局部最优检验*。

为了得到 LMP 检验的一般形式，假设对于每个 $\theta\in\Lambda_1$，P_θ 都具有概率密度 p_θ，即

$$P_D(\tilde{\delta};\theta)=E_\theta\{\tilde{\delta}(Y)\}=\int_\Gamma\tilde{\delta}(y)p_\theta(y)\mu(\mathrm{d}y) \tag{2.5.29}$$

假设 $\{p_\theta(y);\ \theta\in\Lambda_1\}$ 满足正则性要求，则可以交换式(2.5.29)中积分和求导的次序，进而有

$$P_D'(\tilde{\delta};\theta_0)=\int_\Gamma\tilde{\delta}(y)\frac{\partial}{\partial\theta}p_\theta(y)\big|_{\theta=\theta_0}\mu(\mathrm{d}y) \tag{2.5.30}$$

对比式(2.5.30)和式(2.4.4)可知：如果将 $p_1(y)$ 用 $\partial p_\theta(y)/\partial\theta\big|_{\theta=\theta_0}$ 代替，则设计虚警概率为 α 的 LMP 判决准则与设计虚警概率为 α 的尼曼-皮尔逊判决准则相同。基于上述类比，易知式(2.5.26)对应的虚警概率为 α 的 LMP 检验为

$$\tilde{\delta}_{lo}(y)=\begin{cases}1, & \dfrac{\partial}{\partial\theta}p_\theta(y)\big|_{\theta=\theta_0}>\eta p_{\theta_0}(y)\\[2ex] \gamma, & \dfrac{\partial}{\partial\theta}p_\theta(y)\big|_{\theta=\theta_0}=\eta p_{\theta_0}(y)\\[2ex] 0, & \dfrac{\partial}{\partial\theta}p_\theta(y)\big|_{\theta=\theta_0}<\eta p_{\theta_0}(y)\end{cases} \tag{2.5.31}$$

式中，η 和 γ 满足 $P_F(\tilde{\delta}_{lo})=\alpha$。具体推导步骤可以参考 Ferguson 的书(1968)。第 3 章将深入讨论 LMP 检验的特性。

对于上述最优判决准则均无法应用的情况，通常选择将下式与某个门限进行比较来实现判决

$$\frac{\max\limits_{\theta\in\Lambda_1} p_\theta(y)}{\max\limits_{\theta\in\Lambda_0} p_\theta(y)} \tag{2.5.32}$$

其中集合 Λ_0 和 Λ_1 不相交。该检验也称作广义似然检验或最大似然检验，这一检验将在第4章中具体介绍。

2.6 习题

1. 求例 2.2.1 中的二元信道最小贝叶斯风险。

2. Y 是一随机变量，在 H_0 条件下的概率分布函数为

$$p_0(y)=\begin{cases}(2/3)(y+1), & 0\leqslant y\leqslant 1\\ 0, & \text{其他}\end{cases}$$

在 H_1 条件下的概率分布函数为

$$p_1(y)=\begin{cases}1, & 0\leqslant y\leqslant 1\\ 0, & \text{其他}\end{cases}$$

(a) 求在均匀代价和相等先验概率条件下，检测 H_0 和 H_1 的贝叶斯准则和最小贝叶斯风险。

(b) 求均匀代价条件下的极小化极大风险。

(c) 求尼曼-皮尔逊准则和虚警概率 $\alpha\in(0,1)$ 情况下的检测概率。

3. 习题 2 中的 p_j 用下式代替

$$p_j(y)=\frac{(j+1)}{2}e^{-(j+1)|y|},\quad y\in\mathbf{R},\quad j=0,1$$

且习题 2 中问题(a)的先验概率为 $\pi_0=1/4$、$\pi_1=3/4$，问题(a)和问题(b)的代价函数为

$$C_{i,j}=\begin{cases}0, & i=j\\ 1, & i=1,\quad j=0\\ 3/4, & i=0,\quad i=0\end{cases}$$

重做习题 2。

4. 将习题 2 的概率分布函数 p_0 和 p_1 分别改写为

$$p_0(y)=\begin{cases}e^{-y}, & y\geqslant 0\\ 0, & y<0\end{cases},$$

$$p_1(y)=\begin{cases}\sqrt{2/\pi}e^{-y^2/2}, & y\geqslant 0\\ 0, & y<0\end{cases}$$

对于问题(a)，考虑任意先验概率，重做习题 2。

5. 在习题 2 中，假设在 H_0 和 H_1 条件下 Y 的概率密度函数为

$$H_0: p_0(y)=\frac{1}{\sqrt{2\pi}}e^{-y^2/2}, y\in\mathbf{R}$$

$$H_1: p_1(y)=\begin{cases}1/5, & y\in[0,5]\\ 0, & y\notin[0,5]\end{cases}$$

且在习题 2 问题(a)中的先验概率为 $\pi_0=3/4$、$\pi_1=1/4$，重做习题 2。

6. 在习题 2 中，假设对为

$$H_0: Y=N-s$$
$$H_1: Y=N+s$$

其中，s 是一个固定实数且 $s>0$，N 是连续随机变量，其概率密度为

$$p_N(n)=\frac{1}{\pi(1+n^2)},\quad n\in\mathbf{R}$$

重做习题 2。

7. (a) 给出一个假设对

$$H_0: Y=N$$
$$H_1: Y=N+S$$

其中，N 和 S 是独立的随机变量，其概率分布函数为

$$p(x)=\begin{cases}e^{-x}, & x\geqslant 0\\ 0, & x<0\end{cases}$$

求 H_0 和 H_1 之间的最大似然比。

(b) 求采用虚警概率为 α 的尼曼-皮尔逊准则检测问题(a)的检测门限和检测概率。

(c) 考虑假设对

$$H_0: Y_k=N_k,\qquad k=1,\cdots,n$$
$$H_1: Y_k=N_k+S,\quad k=1,\cdots,n$$

其中，$n>1$ 且 N_1，…，N_n 和 S 是独立的随机变量，其概率分布函数在问题(a)中给出，求似然比。

(d) 求采用虚警概率为 α 的尼曼-皮尔逊准则检测问题(c)的检测门限。

8. 证明最小贝叶斯风险函数 V（定义参见 2.3 节）是凹的且在[0，1]区间内连续。（在证明 V 是凹的之后，可以利用[0，1]上的任一凹函数在(0，1)上连续这一事实。）

9. 假设观测序列 Y 是一实数，且二元假设的概率分布函数可以表示为

$$p_0(y)=\begin{cases}1-|y|, & |y|\leqslant 1\\ 0, & |y|>1\end{cases}$$

$$p_1(y)=\begin{cases}(2-|y|)/4, & |y|\leqslant 2\\ 0, & |y|>2\end{cases}$$

(a) 假设代价为

$$C_{01}=2C_{10}>0$$
$$C_{00}=C_{11}=0$$

求解 H_0 和 H_1 的极小化极大检测以及相应的极小化极大风险。

(b) 求解虚警概率为 α 时，H_0 和 H_1 的尼曼-皮尔逊检测以及相应的检测概率。

10. 假设观测一随机变量 Y，且

$$Y=N+\theta\lambda$$

其中，θ 为 1 或 0，λ 是 0～2 之间的一个固定值，N 是一随机变量，在(−1，1)区间满足均匀分布。希望在以下两种假设间进行判决

$$H_0:\theta=0$$
$$H_1:\theta=1$$

(a) 求虚警概率在 0～1 之间的尼曼-皮尔逊判决准则。

(b) 求尼曼-皮尔逊判决准则的检测概率(自变量为虚警概率和参数 λ)，并画出接收机工作特性曲线。

11. 假设观测序列 Y 是一实数，其假设检测问题为

$$H_0:p_0(y)=\exp(-y^2/2)/\sqrt{2\pi},\qquad y\in\mathbf{R}$$
$$H_1:p_1(y)=\exp(-(y-1)^2/2)/\sqrt{2\pi},\ y\in\mathbf{R}$$

代价分配为 $C_{00}=C_{11}=0$，$C_{10}=1$，$C_{01}=N$。求该检测问题在等概率假设条件下的贝叶斯准则和贝叶斯风险，以及当 N 趋于无穷时的极小化极大准则和风险。

12. 考虑单个二元假设问题，其判决准则为 δ，分别用 $P_F(\delta)$ 和 $P_M(\delta)$ 表示虚警概率和漏检概率。性能估计定义为

$$\rho(\delta)\triangleq[P_F(\delta)]^2+[P_M(\delta)]^2$$

δ_0 表示所有随机化判决准则 δ 中使得 $\rho(\delta)$ 最小的一个准则。

(a) 证明 δ_0 一定是似然比检测。

(b) 当 $\pi_0\in[0,1]$时，定义函数 V 为

$$V(\pi_0)=\min_{\delta}[\pi_0P_F+(1-\pi_0)P_M]$$

假设 $\pi_0=1/2$ 时，$V(\pi_0)$在区间[0，1]内达到最大值。证明 δ_0 是先验概率 $\pi_0=1/2$ 时的贝叶斯准则。(提示：$2\rho(\delta)=[P_F(\delta)+P_M(\delta)]^2+[P_F(\delta)-P_M(\delta)]^2$。)

13. 考虑下面的贝叶斯判决问题：实观测 Y 在给定实参数 $\Theta=\theta$ 下的条件概率密度为

$$p_\theta(y)=\begin{cases}\theta e^{-\theta y}, & y\geqslant 0\\ 0, & y<0\end{cases}$$

Θ 是一随机变量，其概率密度为

$$w(\theta)=\begin{cases}\alpha e^{-\alpha\theta}, & \theta\geqslant 0\\ 0, & \theta<0\end{cases}$$

式中，$\alpha>0$。求解下述假设检验的贝叶斯准则和最小贝叶斯风险

$$H_0:\Theta\in(0,\beta)\triangleq\Lambda_0$$
$$H_1:\Theta\in[\beta,\infty)\triangleq\Lambda_1$$

式中，$\beta>0$ 是一个固定值。假设代价函数为

$$C[i,\theta]=\begin{cases}1, & \theta\notin\Lambda_i\\ 0, & \theta\in\Lambda_i\end{cases}$$

14. 将习题 13 中的 Y 看成是由 n 个独立的且具有相同分布的观测值 $Y=Y_1,\cdots,Y_n$ 组成，其条件概率密度与习题 13 相同。重做习题 13，但可不给出贝叶斯风险的闭式解。

15. 考虑混合假设检验问题

$$H_0:Y\text{ 的概率密度}\quad p_0=\frac{1}{2}e^{-|y|},$$
$$y\in\mathbf{R}$$
$$H_1:Y\text{ 的概率密度 }p_\theta=\frac{1}{2}e^{-|y-\theta|},$$
$$y\in\mathbf{R},\quad\theta>0$$

(a) 描述虚警概率为 α 的局部最大势检验并推导它的势函数。

(b) 是否存在一致最大势检验？如果有，求出它并推导它的势函数；如果没有，求解 H_0 和 H_1 的广义似然比检测。

16. 在 2.2 节中，我们求解了二元贝叶斯假设检验问题，试将它推广为 M 元假设问题($M>2$)。

17. 推导 M 元极小化极大假设检验的公式，证明贝叶斯均衡准则(如果存在的话)就是极小化极大准则。

18. 对于 M 元假设，如何得到一个与尼曼-皮尔逊

准则类似的准则？给出求解方法。

19. 以下给出了独立的随机变量序列 Y_1，Y_2，…，Y_n 的一对假设

$$H_0:Y_k \sim \mathcal{N}(\mu_0,\sigma_0^2),\quad k=1,2,\cdots,n$$

$$H_1:Y_k \sim \mathcal{N}(\mu_1,\sigma_1^2),\quad k=1,2,\cdots,n$$

式中，μ_0、μ_1、σ_0^2 和 σ_1^2 均为已知常数。

(a) 证明似然比能表示为关于参数 μ_0、μ_1、σ_0^2 和 σ_1^2，以及总量 $\sum_{k=1}^{n} Y_k^2$ 和 $\sum_{k=1}^{n} Y_k$ 的函数。

(b) 分别描述（$\mu_0=\mu_1$，$\sigma_1^2>\sigma_0^2$）和（$\sigma_0^2=\sigma_1^2$，$\mu_1>\mu_0$）下的尼曼-皮尔逊检验。

(c) 求解 $\mu_0=\mu_1$，$\sigma_1^2>\sigma_0^2$，且 $n=1$ 时的门限值和 ROC。

20. 考虑在 $\mu \triangleq \mu_1>\mu_0=0$ 和 $\sigma \triangleq \sigma_0^2=\sigma_1^2>0$ 的条件下习题 19 的假设检验问题。在 μ 和 σ^2 都已知的情况下，是否存在一致最大势检验？如果有，找到它并证明它是 UMP；如果没有，证明之，并给出广义似然比检验。

21. 假设 $Y=Y_1$，Y_2，…，Y_n 是一组随机观测序列，每个观测以等概率在 0 和 1 中取值。考虑关于 $Y=Y_1$，Y_2，…，Y_n 的二元假设

$$H_0:Y_1,Y_2,\cdots,Y_n \text{ 是独立的}$$

$$H_1:p_1(y_k \mid y_1,y_2,\cdots,y_{k-1}) = \begin{cases} 3/4, & y_k=y_{k-1} \\ 1/4, & y_k \neq y_{k-1},k=2,3,\cdots,n \end{cases}$$

式中，$p_1(y_k \mid y_1,y_2,\cdots,y_{k-1})$ 是 $Y_k=y_k$ 在 $Y_1=y_1$，$Y_2=y_2$，…，$Y_{k-1}=y_{k-1}$ 下的条件概率分布。求均匀代价和先验概率相等条件下的贝叶斯判决准则。

第3章 离散时间信号检测

3.1 引言

第2章讨论了二元假设检验问题中几种基本的最优化准则和设计方法，本章将利用这些方法以及一些相关的方法来推导噪声中的最优信号检测过程。为了避免分析复杂化，本章仅考虑离散时间的信号检测问题，关于连续时间的分析将在第4章给出。由于信号处理通常是以数字运算的方式进行，因此研究离散时间下的信号检测有着非常重要的意义。

3.2节将以第2章为基础，讨论各种信号检测问题的模型并推导出其各自的最优检测器结构与相应的准则集合。对于最优检测器的相对误差概率闭式解不易求解的情况，相应的性能分析方法将在3.3节中具体论述。此外，3.4节和3.5节将介绍三种与第2章不同的信号检测方法，它们分别是序贯、稳健和非参数化检测。

3.2 模型和检测器结构

在基本观测模型中，所观测到的连续时间波形由被加性噪声污染的两个有用信号之一构成，目标是通过处理观测波形中的有限样本，来对这两种可能的信号进行判定。

利用观测空间$(\Gamma, \mathcal{G})=(\mathbf{R}^n, \mathcal{B}^n)$上的假设检验对概念，上述问题可统计建模为

$$\begin{aligned} H_0: Y_k &= N_k + S_{0k}, \quad k = 1,2,\cdots,n \\ H_1: Y_k &= N_k + S_{1k}, \quad k = 1,2,\cdots,n \end{aligned} \tag{3.2.1}$$

式中，$\boldsymbol{Y}=(Y_1, \cdots, Y_n)^{\mathrm{T}}$是由观测波形采样所构成的观测向量，$\boldsymbol{N}=(N_1, \cdots, N_n)^{\mathrm{T}}$是噪声采样向量。$\boldsymbol{S}_0=(S_{01}, \cdots, S_{0n})^{\mathrm{T}}$和$\boldsymbol{S}_1=(S_{11}, \cdots, S_{1n})^{\mathrm{T}}$是两种可能信号的采样[⊖]。这里需要指出，式(3.2.1)中的$\boldsymbol{Y}$不仅可以理解为时间采样向量，还可以表示同时对n个空间分离的传感器或者n组平行滤波器的输出进行采样所产生的向量。鉴于时间模型的结果能等价应用到其他可由式(3.2.1)表示的模型中，因此通常都将下标n看作是时间参数。

如果已知信号和噪声的统计模型，那么可以利用第2章的结论设计H_0和H_1的最优判决过程。从实际应用出发，信号S_0和S_1通常可以分为三类，即完全已知的信号（确定信号）、参数未知（或随机）的确定信号和仅能用概率分布描述的完全随机信号。在一些情况下（例如雷达、声呐系统中），它们中的某个信号（通常是S_0）等于0，因此实际上需要检测淹没在噪声中的信号。对于这种处理方式，通常会假定噪声与每一个假设下的信号都是独立的，同时它的分布概率与假设是否成立无关。尽管在一些特殊应用中噪声可能依赖于

⊖ 本书中如不作特殊说明，向量均指列向量，上标“T”表示转置。

信号(例如在雷达、声呐探测中，噪声有时包含了由地面或其他与目标独立的物体所反射的信号)，但这一假定在大多数应用中都是有效的，因此通常假设噪声部分可用 $\mathbf{R}^n$ 空间上的连续或离散概率密度 p_N 表示。

在上述假设条件下，如果知道$\boldsymbol{S}_j$ 在 $j=0$，1 时的统计分布，就可以计算出式(3.2.1)的似然比。特别地，若给定$\boldsymbol{S}_j=\boldsymbol{s}_j\in\mathbf{R}^n$，则观测数据$\boldsymbol{Y}$的条件概率密度为(在 H_j 下)

$$p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{s}_j),\quad \boldsymbol{y}\in\mathbf{R}^n \tag{3.2.2}$$

由式(3.2.2)可以得到 Y 在 H_j 下的概率密度为

$$p_j(\boldsymbol{y})=E\{p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{S}_j)\},\quad \boldsymbol{y}\in\mathbf{R}^n \tag{3.2.3}$$

需要注意的是，上式是关于信号S_j 的期望。进而可将似然比写作

$$L(\boldsymbol{y})=\frac{E\{p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{S}_1)\}}{E\{p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{S}_0)\}},\quad \boldsymbol{y}\in\mathbf{R}^n \tag{3.2.4}$$

因此，通过计算式(3.2.4)即可得到式(3.2.1)的最优判决过程，而通过计算相应判决区域的概率可对其性能进行分析。下面将结合典型模型对这类判决问题展开讨论。

模型 3.2.1　独立噪声中的确定信号检测

在许多问题中，信号$\boldsymbol{S}_0$ 和$\boldsymbol{S}_1$ 是完全确定的。特别地，$\boldsymbol{S}_j=\boldsymbol{s}_j$，$\boldsymbol{s}_j\in\mathbf{R}^n$ 通常在设计时就已知。该问题有时也称作相干检测问题。在这一情况下，式(3.2.4)中的 $L(\boldsymbol{y})$变为

$$L(\boldsymbol{y})=\frac{p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{s}_1)}{p_{\boldsymbol{N}}(\boldsymbol{y}-\boldsymbol{s}_0)}=\frac{p_{\boldsymbol{N}}(y_1-s_{11},\cdots,y_n-s_{1n})}{p_{\boldsymbol{N}}(y_1-s_{01},\cdots,y_n-s_{0n})} \tag{3.2.5}$$

从上式可以发现，如果 $p_{\boldsymbol{N}}$ 已知，就很容易确定最优检测器的结构。不过，一般而言，式(3.2.5)是观测数据的复杂函数，往往难以应用和分析，特别是在样本数目(n)较大时更为困难。因此需要对概率分布函数作进一步的简化，否则很难得到基于式(3.2.5)的判决器的更多结论。例如，对于极小化极大值或尼曼-皮尔逊判决准则的阈值设定，计算检测器的性能需要 n 重积分 $\int_{\{L(\boldsymbol{y})>\tau\}}p_j(\boldsymbol{y})\mu(\mathrm{d}\boldsymbol{y})$，所涉及的运算量十分巨大。噪声样本 N_1，…，N_n 统计独立是式(3.2.1)在已知信号下的一个特殊情况。在这种情况下，有

$$p_{\boldsymbol{N}}(\boldsymbol{y})=\prod_{k=1}^{n}p_{N_k}(y_k) \tag{3.2.6}$$

式中，p_{N_k}是 N_k 的边缘密度，$L(\boldsymbol{y})$可表示为

$$L(\boldsymbol{y})=\prod_{k=1}^{n}L_k(y_k) \tag{3.2.7}$$

式中，$L_k(y_k)=p_{N_k}(y_k-s_{1k})/p_{N_k}(y_k-s_{0k})$。由于 $\log(x)$是 x 的严格递增函数，因此将 $L(\boldsymbol{y})$与阈值 τ 进行比较和将 $\log L(\boldsymbol{y})$与阈值 $\log(\tau)$进行比较等价。因此，最优判决可写为

$$\tilde{\delta}_{\mathrm{o}}(\boldsymbol{y})=\begin{cases}1, & \sum\limits_{k=1}^{n}\log L_k(y_k)>\log\tau\\ \gamma, & \sum\limits_{k=1}^{n}\log L_k(y_k)=\log\tau\\ 0, & \sum\limits_{k=1}^{n}\log L_k(y_k)<\log\tau\end{cases} \tag{3.2.8}$$

如图 3.2.1 所示，该最优判决结构由非线性瞬态时变 $\log L_k$、累加器以及累加器后的阈值比较器构成。下面两个例子将对式(3.2.8)在实际中的两种特殊情况进行说明。

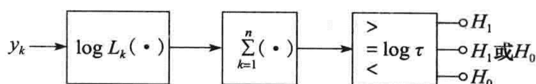


图 3.2.1　独立噪声情况下的相干信号检测结构

例 3.2.1　独立同分布高斯噪声下的相干检测

假设噪声样本 N_1，…，N_n 满足独立同分布，且边缘分布服从 $N(0, \sigma^2)$。例如在通信系统的接收端，噪声主要来自接收电路中电子运动产生的热噪声。为了方便起见，假设 $\boldsymbol{s}_0=\boldsymbol{0}$，其中 $\boldsymbol{0}$ 表示 n 维 0 向量，$\boldsymbol{s}$ 表示 $\boldsymbol{s}_1$(注意，该假设不会造成任何损失，因为在重新定义观测数据 $\boldsymbol{y}'=\boldsymbol{y}-\boldsymbol{s}_0$ 后，信号在 H_0 下变为 $\boldsymbol{0}$，在 H_1 下变为 $\boldsymbol{s}=\boldsymbol{s}_1-\boldsymbol{s}_0$)。因此，有 $\log L_k(y_k)=s_k(y_k-s_k/2)/\sigma^2$(见例 2.2.2)，则最优判决为

$$\tilde{\delta}_o(\boldsymbol{y})=\begin{cases}1, & \sum_{k=1}^{n}s_k\left(y_k-\frac{s_k}{2}\right)>\tau' \\ \gamma, & \sum_{k=1}^{n}s_k\left(y_k-\frac{s_k}{2}\right)=\tau' \\ 0, & \sum_{k=1}^{n}s_k\left(y_k-\frac{s_k}{2}\right)<\tau'\end{cases} \tag{3.2.9}$$

式中，$\tau'\triangleq\sigma^2\log\tau$。其结构如图 3.2.2a 所示。注意，$-\frac{1}{2}\sum_{k=1}^{n}s_k^2$ 可以合并到阈值中，因此与式(3.2.9)等价的判决是将 $\sum_{k=1}^{n}s_ky_k$ 同阈值 $\tau''=\tau'+\frac{1}{2}\sum_{k=1}^{n}s_k^2$ 进行比较。此结构称为相关检测器(简称相关器)，如图 3.2.2b 所示。

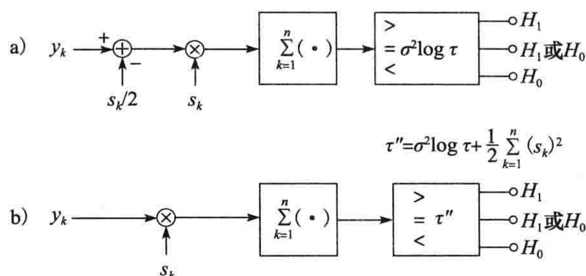


图 3.2.2　独立同分布高斯噪声中相干信号的最优检测

通过比较线性系统的输出和阈值来检测信号，是最优检测器在高斯噪声中的一个重要特征。特别地，有 $\sum_{k=1}^{n}s_ky_k=\sum_{k=-\infty}^{n}h_{n-k}y_k$，其中

$$h_k=\begin{cases}s_{n-k}, & 0\leqslant k\leqslant n-1 \\ 0, & \text{其他}\end{cases} \tag{3.2.10}$$

相应地，检测器可以看作是将观测序列 y_1，…，y_n 输入一个数字线性滤波器，然后在时间 n 对输出进行采样并与阈值进行比较的系统。这种检测结构也称为匹配滤波器。

例 3.2.2　独立同分布拉普拉斯噪声下的相干检测

如例 3.2.1 所示，假设噪声样本 N_1，…，N_n 是独立同分布的，但边缘概率服从拉普拉斯分布，即

$$p_{N_k}(y_k)=\frac{\alpha}{2}e^{-\alpha|y_k|},\quad y_k\in\mathbf{R} \tag{3.2.11}$$

式中，$\alpha>0$ 是密度的一个标量参数。该模型一般用于描述通信接收机的脉冲噪声。与高斯模型相比，它具有较长的“拖尾”，也就意味着大的观测值有更高的概率。

式(3.2.11)中的函数 $\log L_k(y_k)$ 可由 $L_k(y_k)=\alpha(|y_k|-|y_k-s_k|)$ 得到，进而有下面的表达式成立

$$\log L_k(y_k)=\begin{cases}-\alpha|s_k|, & \operatorname{sgn}(s_k)y_k\leqslant 0\\ \alpha|2y_k-s_k|, & 0<\operatorname{sgn}(s_k)y_k<|s_k|\\ +\alpha|s_k|, & \operatorname{sgn}(s_k)y_k\geqslant|s_k|\end{cases} \tag{3.2.12}$$

式中，sgn 表示符号函数

$$\operatorname{sgn}(x)=\begin{cases}+1, & x>0\\ 0, & x=0\\ -1, & x<0\end{cases} \tag{3.2.13}$$

在图 3.2.3 中，给出了函数 $\log L_k(y_k)$ 在两种情况($s_k<0$ 和 $s_k>0$)下的具体描述。观察图 3.2.3可以发现，基于 $\sum_{k=1}^{n}\log L_k(y_k)$ 的最优检测器能够通过下式实现

$$\tilde{\delta}_o(\mathbf{y})=\begin{cases}1, & \sum_{k=1}^{n}\operatorname{sgn}(s_k)l_k(y_k)>\tau'\\ \gamma, & \sum_{k=1}^{n}\operatorname{sgn}(s_k)l_k(y_k)=\tau'\\ 0, & \sum_{k=1}^{n}\operatorname{sgn}(s_k)l_k(y_k)<\tau'\end{cases} \tag{3.2.14}$$

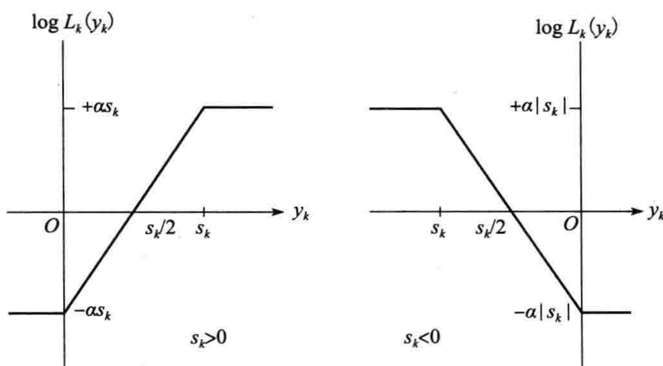


图 3.2.3　拉普拉斯噪声中相干检测的单个样本对数似然比

其中 $\tau'\triangleq\frac{\log\tau}{2\alpha}$，且函数 l_k 具有如下形式

$$l_k(x)=\begin{cases}-|s_k|/2, & x\leqslant -|s_k|/2\\ x, & -|s_k|/2<x<|s_k|/2\\ +|s_k|/2, & x\geqslant +|s_k|/2\end{cases}\tag{3.2.15}$$

注意，如果 $x\in(-|s_k|/2, +|s_k|/2)$，则 l_k 是线性的，否则 $l_k=|s_k|\operatorname{sgn}(x)/2$。该函数通常称作软限幅器或放大限幅器。式(3.2.14)的系统表述如图 3.2.4 所示。

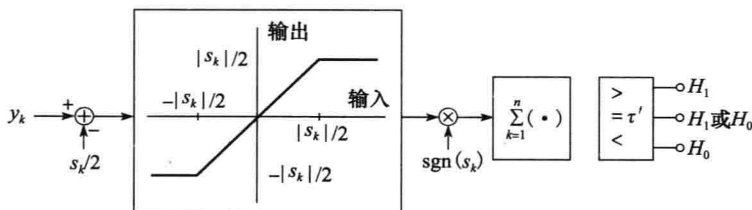


图 3.2.4　拉普拉斯噪声下相干信号的最优检测

对比图 3.2.2a 和图 3.2.4 的检测器结构可以发现：两个系统都是通过将每个 y_k 减去 $s_k/2$来“中心化”观察数据。图 3.2.2a 中的系统把“中心化”后的数据与已知信号作相关，并把相关的输出与阈值相比较。与之不同的是，图 3.2.4 所描述的检测器先对“中心化”后的数据进行软限幅，然后对这些软限幅后的观测数据和信号的符号序列作相关。软限幅的作用就是减小大观测值对求和的影响，从而能够比图 3.2.2a 所示的线性系统容纳更大的噪声值。与之前讨论的高斯噪声相比，图 3.2.4 所描述的处理过程结合了拉普拉斯随机变量的特性。◀

例 3.2.3　独立同分布噪声下相干信号的局部最优检测

在许多检测问题中，接收信号的形式是已知的，但其幅度未知。为了对这类问题进行建模，考虑复合假设检验问题如下

$$\begin{aligned}H_0&:Y_k=N_k, & k&=1,2,\cdots,n\\ H_1&:Y_k=N_k+\theta s_k, & k&=1,2,\cdots,n,\quad \theta>0\end{aligned}\tag{3.2.16}$$

式中，$\mathbf{s}=(s_1, \cdots, s_n)^{\mathrm{T}}$ 是已知的信号，$\mathbf{N}=(N_1, \cdots, N_n)^{\mathrm{T}}$ 是一个由独立同分布分量组成的连续随机噪声向量，其边缘概率密度函数为 p_{N_k}，θ 是信号强度参数。对于给定的 θ，H_0 和 H_1 的似然比可以写作

$$L_\theta(\mathbf{y})=\prod_{k=1}^{n}\frac{p_{N_k}(y_k-\theta_{sk})}{p_{N_k}(y_k)}\tag{3.2.17}$$

注意，除非遇到特殊情况(例 3.2.1 中的高斯噪声就是一种特殊情况)，否则判决区域 $\Gamma_\theta=\{\mathbf{y}\in\mathbf{R}^n\,|\,L_\theta(\mathbf{y})>\tau\}$一般依赖于 θ。因此，式(3.2.16)的 UMP(一致最大势)假设检验只在特定噪声模型下存在。然而，式(3.2.16)的 LMP 假设检验却有着特别简单且直观合理的结构，因此考虑这种情况下的局部最优检测是非常有意义的。

假设边缘概率密度函数 p_{N_k} 足够光滑(存在微分)，H_0 对于 H_1 的局部最优检测如下[见式(2.2.31)]

$$\tilde{\delta}_{\mathrm{lo}}(\mathbf{y})=\begin{cases}1, & \dfrac{\partial}{\partial\theta}L_\theta(\mathbf{y})\,|_{\theta=0}>\tau\\ \gamma, & \dfrac{\partial}{\partial\theta}L_\theta(\mathbf{y})\,|_{\theta=0}=\tau\\ 0, & \dfrac{\partial}{\partial\theta}L_\theta(\mathbf{y})\,|_{\theta=0}<\tau\end{cases}\tag{3.2.18}$$

对式(3.2.17)取偏导数，得到

$$\frac{\partial}{\partial\theta}L_\theta(\mathbf{y})\,|_{\theta=0}=\sum_{k-1}^{n}s_k g_{\mathrm{lo}}(y_k)\tag{3.2.19}$$

其中，$g_{\mathrm{lo}}(x)\triangleq-p'_{N_1}(x)/p_{N_1}(x)$，$p'_{N_1}(x)=\mathrm{d}p_{N_1}(x)/\mathrm{d}x$。此结构如图 3.2.5 所示，它由无记忆的非线性因子 g_{lo}和相关器组成，也称为非线性相关器。

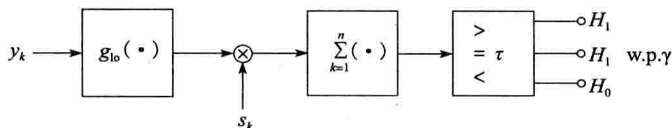


图 3.2.5 独立同分布噪声中相干信号的局部最优检测器

与似然比类似，局部最优非线性因子 g_{lo}对观察数据进行一定的处理，从而尽可能减少噪声带来的不利影响。例如，对于服从$\mathcal{N}(0,\sigma^2)$分布的噪声，有 $g_{\mathrm{lo}}(x)=x/\sigma^2$，此时图 3.2.5仅是图 3.2.2b 中的相关检测器(这是必然成立的，因为该检测器是 UMP 检测器，所以它也一定是 LMP 检测器)。当拉普拉斯噪声密度为式(3.2.11)时，有 $g_{\mathrm{lo}}(x)=\alpha\mathrm{sgn}(x)$，则局部最优检测器是对信号与观察数据中信号的符号序列作相关，如图 3.2.6 所示。在这种情况下，函数 $g_{\mathrm{lo}}(x)$称为硬限幅器。对于用柯西概率密度描述的噪声而言，其概率密度为

$$p_{N_1}(x)=\frac{1}{\pi(1+x^2)},\quad -\infty<x<+\infty\tag{3.2.20}$$

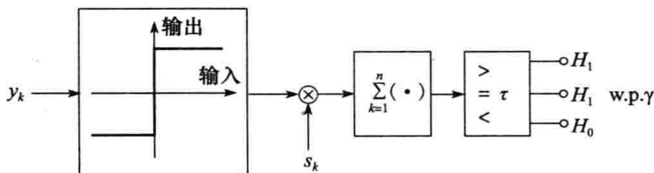


图 3.2.6 拉普拉斯噪声中的局部最优检测器

它具有比拉普拉斯噪声更严重的拖尾。此时，$g_{\mathrm{lo}}(x)=2x/(1+x^2)$在 $x=0$ 附近近似线性，之后再次下降(渐近地)至零，即 $\lim\limits_{x\to\pm\infty}g_{\mathrm{lo}}(x)=0$。显然，检测器忽略了幅度非常大的观测值。图 3.2.7 给出了该检测器的基本结构。阻塞噪声是对上述非线性的一种近似，具体为

$$g(x)=\begin{cases}x, & |x|\leqslant K\\ 0, & |x|>K\end{cases}\tag{3.2.21}$$

式中，k 为大于零的常数，常用来消除特别强的脉冲噪声。 ◀

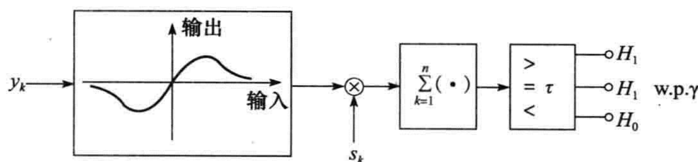


图 3.2.7　柯西噪声中的局部最优检测器

模型 3.2.2　高斯噪声中的确定信号检测

如果式(3.2.1)中的噪声样本 N_k 相互不独立，即使在确知信号情况下，式(3.2.4)得到的最优解也不具有式(3.2.5)以外的任何特殊结构。不过，如果信号是确定的($\boldsymbol{S}_j=\boldsymbol{s}_j$)并且噪声向量 $\boldsymbol{N}$ 服从多变量高斯分布，则最优检测器具有非常简单且易于实现的结构，同时能分析与之相关的检测性能。此外，高斯噪声的假设在实际应用中比较合理，与之对应的最优系统即使用在非高斯噪声情况下也能具有较好的性能。下面将对这一特殊情况进行详细讨论。

假设信号 $\boldsymbol{S}_j$ 的取值已知且 $\boldsymbol{s}_j\in\mathbf{R}^n$，噪声向量 $\boldsymbol{N}$ 是均值为零、方差为 $\boldsymbol{\Sigma}_N$ 的高斯随机向量。(由于通常可以从 $\boldsymbol{y}$ 中减去一个非零噪声均值来产生新的零均值噪声的观测数据，因此零均值噪声假设不影响结果的一般性)。

对于均值向量为 $\mu\triangleq E\{\boldsymbol{X}\}\in\mathbf{R}^n$ 和 $n\times n$ 协方差矩阵为 $\boldsymbol{\Sigma}\triangleq E\{(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^{\mathrm{T}}\}$ 的高斯随机向量，其概率密度函数为

$$p_{\boldsymbol{x}}(\boldsymbol{x})=\frac{1}{(2\pi)^{n/2}\,|\boldsymbol{\Sigma}|^{1/2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\},\quad \boldsymbol{x}\in\mathbf{R}^n \tag{3.2.22}$$

式中，$|\boldsymbol{\Sigma}|$ 表示 $\boldsymbol{\Sigma}$ 的行列式，$\boldsymbol{\Sigma}^{-1}$ 表示 $\boldsymbol{\Sigma}$ 的逆。该随机变量服从 $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ 分布。注意，协方差矩阵总是非负定的(对于所有 $\boldsymbol{x}\in\mathbf{R}^n$，$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{x}\geqslant 0$)。此外，在式(3.2.22)中，假设 $\boldsymbol{\Sigma}$ 是正定的($\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{x}>0$，对于除 $\boldsymbol{x}=0$ 的任意 $\boldsymbol{x}\in\mathbf{R}^n$)。$\boldsymbol{\Sigma}$ 的正定意味着 $|\boldsymbol{\Sigma}|>0$ 且 $\boldsymbol{\Sigma}^{-1}$ 存在。如果 $\boldsymbol{\Sigma}$ 非正定，则意味着 $\boldsymbol{X}$ 至少有一个分量能由其他分量线性表示，即存在冗余。除非特别说明，否则在处理高斯随机向量时均假设协方差矩阵是正定的。关于高斯随机向量的分析讨论，读者可以参考 Thomas(1986)的书。

利用之前给定的假设 $\boldsymbol{S}_j$ 和 $\boldsymbol{N}$，可以直接计算出其似然比。因为信号是已知的，它对观察数据 $\boldsymbol{Y}$ 的分布的影响只是在 $\boldsymbol{N}$ 的分布上增加了一个均值，所以在 H_j 假设下($j=0$ 或 1)，$\boldsymbol{Y}$ 服从 $\mathcal{N}(\boldsymbol{s}_j,\boldsymbol{\Sigma}_N)$ 分布。可以得出似然比如下

$$\begin{aligned}
L(\boldsymbol{y})&=\frac{p_1(\boldsymbol{y})}{p_0(\boldsymbol{y})}\\
&=\frac{\dfrac{1}{(2\pi)^{n/2}\,|\boldsymbol{\Sigma}_N|^{1/2}}\exp\left\{-\dfrac{1}{2}(\boldsymbol{y}-\boldsymbol{s}_1)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{y}-\boldsymbol{s}_1)\right\}}{\dfrac{1}{(2\pi)^{n/2}\,|\boldsymbol{\Sigma}_N|^{1/2}}\exp\left\{-\dfrac{1}{2}(\boldsymbol{y}-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{y}-\boldsymbol{s}_0)\right\}}\\
&=\exp\left\{\boldsymbol{s}_1^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{y}-\boldsymbol{s}_0^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{y}-\frac{1}{2}\boldsymbol{s}_1^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{s}_1+\frac{1}{2}\boldsymbol{s}_0^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{s}_0\right\}\\
&=\exp\left\{(\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\left(\boldsymbol{y}-\frac{\boldsymbol{s}_0+\boldsymbol{s}_1}{2}\right)\right\},\quad \boldsymbol{y}\in\mathbf{R}^n
\end{aligned} \tag{3.2.23}$$

其中，在倒数第二个等式中，利用 $\boldsymbol{\Sigma}_N$(因为 $\boldsymbol{\Sigma}_N^{-1}$)是对称矩阵的性质，从而有 $\boldsymbol{s}_j^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{y}=\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{s}_j$。

对比式(3.2.23)与对应的标量形式(式(2.2.27))可以发现：此处考虑的问题是例2.2.2中高斯误差下位置检验的向量形式，其中把μ_0和μ_1替换为向量$\boldsymbol{s}_0$和$\boldsymbol{s}_1$，而把噪声方差σ^2替换为噪声协方差矩阵$\boldsymbol{\Sigma}_N$。

基于式(3.2.23)的最优检测形式可通过求对数进行简化，同时由于$\frac{1}{2}(\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{s}_0+\boldsymbol{s}_1)$不依赖于$\boldsymbol{y}$，所以可被合并到判断阈值中。因此，最优判决可表示为

$$\widetilde{\delta}_{\mathrm{o}}(\boldsymbol{y})=\begin{cases}1, & (\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{y}>\tau' \\ \gamma, & (\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{y}=\tau' \\ 0, & (\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{y}<\tau'\end{cases} \tag{3.2.24}$$

式中，$\tau'=\log\tau+\frac{1}{2}(\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{s}_0+\boldsymbol{s}_1)$。可以得到

$$(\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{y}=\widetilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{y}=\sum_{k=1}^{n}\widetilde{s}_k y_k \tag{3.2.25}$$

式中，$\widetilde{\boldsymbol{s}}=\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{s}_1-\boldsymbol{s}_0)$。因此，当实际信号$\boldsymbol{s}$用“伪信号”$\widetilde{\boldsymbol{s}}$代替后，该检测器结构与图3.2.2b中的相关检测器完全相同。所以，对于高斯情况，相关噪声下的应用不会比独立噪声更困难。

下面将对式(3.2.24)所给出的检测器结构作进一步分析。首先考虑式(3.2.24)的检测性能。检测量$T(\boldsymbol{Y})\triangleq(\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{Y}$是高斯随机向量$\boldsymbol{Y}$的线性变换。多元高斯分布的一个基本性质是高斯向量的线性变换仍然是高斯的。在这种情况下，该变换是高斯向量到$\mathbf{R}$的变换，$T(\boldsymbol{Y})$是一个高斯随机变量，因此可以通过求取其在两种假设下的均值和方差来完全描述它在H_0和H_1下的分布特性。在H_j下，$T(\boldsymbol{Y})$的均值由下式给出

$$\begin{aligned}E\{T(\boldsymbol{Y})\mid H_j\}&=E\{\widetilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{Y}\mid H_j\}=\widetilde{\boldsymbol{s}}^{\mathrm{T}}E\{\boldsymbol{Y}\mid H_j\}\\&=\widetilde{\boldsymbol{s}}^{\mathrm{T}}E\{\boldsymbol{N}\mid H_j\}+\widetilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{s}_j=\widetilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{s}_j\triangleq\widetilde{\mu}_j\end{aligned} \tag{3.2.26}$$

式中，$\widetilde{\boldsymbol{s}}=\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{s}_1-\boldsymbol{s}_0)$。类似地，$T(\boldsymbol{Y})$在$H_j$下的方差为

$$\begin{aligned}\operatorname{var}\{T(\boldsymbol{Y})\mid H_j\}&=E\{(\widetilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{Y}-\widetilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{s}_j)^2\mid H_j\}=E\{(\widetilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{N})^2\}\\&=E\{\widetilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{N}\boldsymbol{N}^{\mathrm{T}}\widetilde{\boldsymbol{s}}\}=\widetilde{\boldsymbol{s}}^{\mathrm{T}}E\{\boldsymbol{N}\boldsymbol{N}^{\mathrm{T}}\}\widetilde{\boldsymbol{s}}\\&=\widetilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{\Sigma}_N\widetilde{\boldsymbol{s}}=(\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{s}_1-\boldsymbol{s}_0)\triangleq d^2\end{aligned} \tag{3.2.27}$$

注意，$T(\boldsymbol{Y})$的方差与假设无关。此外，$\boldsymbol{\Sigma}_N$正定意味着$\boldsymbol{\Sigma}_N^{-1}$仍是正定矩阵，即除非$\boldsymbol{s}_1=\boldsymbol{s}_0$，否则$d^2>0$。

通过上述分析，可以在H_j下得到$T(\boldsymbol{Y})\sim\mathcal{N}(\widetilde{\mu}_j, d^2)$。该结论表明，不考虑其他因素时，式(3.2.24)中的随机分布γ是不相关的。因此，在H_j假设下选取H_1的概率如下

$$P_j(\Gamma_1)=\frac{1}{\sqrt{2\pi}d}\int_{\tau'}^{\infty}\mathrm{e}^{-(x-\widetilde{\mu}_j)^2/2d^2}\,\mathrm{d}x=1-\Phi\left(\frac{\tau'-\widetilde{\mu}_j}{d}\right) \tag{3.2.28}$$

式中，τ'由式(3.2.24)给出，d是d^2的正根。针对贝叶斯问题，则可以将式(3.2.28)写为原始阈值τ的表达式来简化表示，即

$$P_j(\Gamma_1)=\begin{cases}1-\Phi\left(\dfrac{\log\tau}{d}+\dfrac{d}{2}\right), & j=0\\[2ex] 1-\Phi\left(\dfrac{\log\tau}{d}-\dfrac{d}{2}\right), & j=1\end{cases} \tag{3.2.29}$$

对比式(3.2.29)和式(2.2.31)可以看出，该问题的贝叶斯性能与例 2.2.2 中简单位置检验的性能相同，式(2.2.31)中的 d 就是式(3.2.27)中 $n=1$ 时的 d。类似地，极小化极大风险假设检验以及尼曼-皮尔逊假设检验的性能和阈值选取也与例 2.3.1 以及例 2.4.1 的标量情况相同，即将式(3.2.28)中的$\tilde{\mu}_j$ 和 d 等同于式(3.2.30)中的 μ 和 σ。例如，对于显著性水平(或虚警概率)为 α 的尼曼-皮尔逊检测，可通过假设 $P_F(\tilde{\delta}_0)=P_0(\Gamma_1)=\alpha$ 来得到阈值

$$\tau' = d\Phi^{-1}(1-\alpha)+\tilde{\mu}_0 \tag{3.2.30}$$

其对应的检测概率为

$$P_D(\tilde{\delta}_{NP}) = P_1(\Gamma_1) = 1-\Phi\left(\frac{\tau'-\tilde{\mu}_1}{d}\right) = 1-\Phi(\Phi^{-1}(1-\alpha)-d) \tag{3.2.31}$$

对比式(3.2.31)和式(2.4.16)，一般情况下，该问题的检测概率曲线和 ROC 即为图 2.4.3和图 2.4.4。

接下来的备注将讨论高斯噪声下检测相干信号的一些特性。

备注 3.2.1　d^2 详解

在前面的讨论中，我们知道高斯噪声的最优确定性信号检测性能一般随 d 的增加而单调提高。正如例 2.4.2 所示：d 的取值(或者用其平方更合适)可以用来衡量信噪比。为了说明这一点，先考虑在例 3.2.1 中所讨论的情况。其中信号 $\boldsymbol{s}_0=0$，$\boldsymbol{s}_1=\boldsymbol{s}$，独立同分布噪声服从$\mathcal{N}(0,\sigma^2)$分布，对应 $\boldsymbol{\Sigma}_N=\sigma^2\boldsymbol{I}$ 的多变量高斯分布，$\boldsymbol{I}$ 表示 $n\times n$ 单位矩阵，即

$$\boldsymbol{\Sigma}_N = \begin{pmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & \sigma^2 & 0 \\ 0 & \cdots & 0 & 0 & \sigma^2 \end{pmatrix} \tag{3.2.32}$$

此时 $\boldsymbol{\Sigma}_N^{-1}=\sigma^{-2}\boldsymbol{I}$，则 d^2 为

$$d^2 = (\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{s}_1-\boldsymbol{s}_0) = \frac{\boldsymbol{s}^{\mathrm{T}}\boldsymbol{I}\boldsymbol{s}}{\sigma^2} = \frac{\boldsymbol{s}^{\mathrm{T}}\boldsymbol{s}}{\sigma^2} = \frac{1}{\sigma^2}\sum_{k=1}^{n}s_k^2 = n\,\frac{\overline{s^2}}{\sigma^2} \tag{3.2.33}$$

式中，$\overline{s^2}\triangleq\frac{1}{n}\sum_{k=1}^{n}s_k^2$ 为信号的平均能量。注意，$\sigma^2\triangleq\frac{1}{n}\sum_{k=1}^{n}E\{N_k^2\}$ 是噪声的平均能量，而 d^2 在此处由信号和噪声的平均能量比乘以样本数给出。因此，增大它们中任何一个的取值都可以提升检测器性能，且随着它们任一个值的增加，都能达到最优性能。

在非独立同分布和 $\boldsymbol{s}_0=0$ 及 $\boldsymbol{s}_1=\boldsymbol{s}$ 的情况下，同样可得到有关 d^2 的类似表述。特别地，在式(3.2.10)中，可以将 $\sum_{k=1}^{n}\tilde{s}_k y_k$ 视为线性时不变滤波器在 n 时刻的输出，其中滤波器的脉冲响应为

$$\tilde{h}_k = \begin{cases} \tilde{s}_{n-k}, & 0\leqslant k\leqslant n-1 \\ 0, & \text{其他} \end{cases} \tag{3.2.34}$$

如果该滤波器的输入只由信号 s_1，…，s_n 组成，则 n 时刻的输出为

$$\sum_{k=1}^{n}\tilde{s}_k s_k = \boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{s} = d^2$$

因此，在采样时刻与信号对应的输出能量为 d^4。另一方面，如果滤波器输入只为噪声，则 n 时刻的输出为随机量 $\sum_{k=1}^{n}\widetilde{s}_k N_k$，对应的能量为

$$E\left\{\left(\sum_{k=1}^{n}\widetilde{s}_k N_k\right)^2\right\}=E\{(\widetilde{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{N})^2\}=d^2 \tag{3.2.35}$$

上式中利用了式(3.2.27)。对于 $\widetilde{h}_k$ 来说，信号和噪声分别对应的输出能量之比可表示为

$$\frac{\left(\sum_{k=1}^{n}\widetilde{s}_k s_k\right)^2}{E\left\{\left(\sum_{k=1}^{n}\widetilde{s}_k N_k\right)^2\right\}}=\frac{d^4}{d^2}=d^2 \tag{3.2.36}$$

显然，在一般意义上，d^2 的取值可理解为最优检测器所决定的滤波器在采样时刻 n 的输出信噪比。输出 SNR 越高，通过比较样本输出与阈值来检测信号的性能就越好，这也与之前所述检测性能是 d^2 的单调递增函数相吻合。

注意，在 n 时刻，式(3.2.34)描述的滤波器 h_k 的输出信噪比是所有脉冲响应长度为 n 的线性滤波器的最大值(参见习题 1)。由于该结果只依赖于 $\boldsymbol{N}$ 的零均值以及方差为 $\boldsymbol{\Sigma}_N$，因此即使非高斯噪声情况下该结论依然有效。然而，式(3.2.34)的最优性在贝叶斯假设检验、极小化极大风险假设检验与尼曼-皮尔逊假设检验下与高斯噪声的假设密切相关。

对于一般信号来说，独立同分布情况下的 d^2 取值也有另一种解释。此时，有

$$d^2=\frac{1}{\sigma^2}\|\boldsymbol{s}_1-\boldsymbol{s}_0\|^2 \tag{3.2.37}$$

其中 $\|\boldsymbol{s}_1-\boldsymbol{s}_0\|$ 表示信号向量 $\boldsymbol{s}_0$ 和 $\boldsymbol{s}_1$ 的欧氏距离

$$\|\boldsymbol{s}_1-\boldsymbol{s}_0\|=\left[\sum_{k=1}^{n}(s_{1k}-s_{0k})^2\right]^{1/2}$$

从上式中可以发现：信号向量距离越远，得到的检测性能就会越好。在非独立同分布噪声情况下，我们同样可以得到类似的结论，具体分析将在下节中论述。

备注 3.2.2 简化为独立同分布噪声情况

由于 $\boldsymbol{\Sigma}_N$ 是 $n\times n$ 的对称正定矩阵，可以用它的结构特性来进一步认识最优检测系统。一个 $n\times n$ 矩阵 $\boldsymbol{\Sigma}_N$ 的特征值 λ_1，…，λ_n 和其对应的特征向量 $\boldsymbol{v}_1$，…，$\boldsymbol{v}_n$ 是满足 $\boldsymbol{\Sigma}_N\boldsymbol{v}_k=\lambda_k\boldsymbol{v}_k$ 的解(矩阵的特征值是唯一的，但特征向量不唯一)。由于 $\boldsymbol{\Sigma}_N$ 为对称的正定矩阵，其特征值为正实数，且特征向量的选取可满足正交性(对于所有的 l，$k=1$，…，n，如果 $k\neq l$，有 $\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{v}_l=0$ 和 $\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{v}_k=1$ 成立)。进而可以将 $\boldsymbol{\Sigma}_N$ 写为

$$\boldsymbol{\Sigma}_N=\sum_{k=1}^{n}\lambda_k\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}} \tag{3.2.38}$$

式(3.2.38)称为 $\boldsymbol{\Sigma}_N$ 的谱分解，其有效性可以简单地通过正交的特征向量集合 $\boldsymbol{v}_1$，…，$\boldsymbol{v}_n$ 构成空间 $\mathbf{R}^n$ 的一组基来证明(正如在 $\mathbf{R}^n$ 中用 n 个线性不相关向量来表示任意的 $x\in\mathbf{R}^n$ ㊀)。注

㊀ 事实上，对于任意的 $x\in\mathbf{R}^n$，都有 $\boldsymbol{x}=\sum_{k=1}^{n}c_k\boldsymbol{v}_k$ 和 $c_k=\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{x}$，因此可以得到

$$\boldsymbol{\Sigma}_N\boldsymbol{x}=\sum_{k=1}^{n}c_k\boldsymbol{\Sigma}_N\boldsymbol{v}_k^{\mathrm{T}}=\sum_{k=1}^{n}\lambda_k\boldsymbol{v}_k c_k=\sum_{k=1}^{n}\lambda_k\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{x}=\left(\sum_{k=1}^{n}\lambda_k\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}\right)\boldsymbol{x}$$

意，矩阵 $\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}$ 乘以一个向量 $\boldsymbol{x}$，可以得到 $\boldsymbol{x}$ 在 $\boldsymbol{v}$ 方向的投影。利用式(3.2.38)，可直接得到 $\boldsymbol{\Sigma}_N^{-1}=\sum_{k=1}^{n}\lambda_k^{-1}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}$，进而可将最优检测器的统计量 $T(\boldsymbol{y})$表示为

$$T(\boldsymbol{y})=(\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{y}=\sum_{k=1}^{n}(\hat{s}_{1k}-\hat{s}_{0k})\,\hat{y}_k \tag{3.2.39}$$

其中

$$\hat{y}_k=\boldsymbol{v}_k^{\mathrm{T}}y/\sqrt{\lambda_k},\quad k=1,\cdots,n \tag{3.2.40}$$

和

$$\hat{s}_{jk}=\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{s}_j/\sqrt{\lambda_k},\quad k=1,\cdots,n,\quad j=0,1$$

注意，$\boldsymbol{y}$ 可以由$\hat{\boldsymbol{y}}$通过 $\boldsymbol{y}=\sum_{k=1}^{n}\sqrt{\lambda_k}\,\hat{y}_k\boldsymbol{v}_k$ 得到，因此$\hat{\boldsymbol{y}}$与 $\boldsymbol{y}$ 是等价的观测数据。对$\boldsymbol{Y}$($\hat{\boldsymbol{y}}$对应的随机向量)来说，其关于式(3.2.1)的假设检验对可以写为

$$\begin{aligned}H_0:\hat{\boldsymbol{Y}}&=\hat{\boldsymbol{N}}+\hat{s}_0\\H_1:\hat{\boldsymbol{Y}}&=\hat{\boldsymbol{N}}+\hat{s}_1\end{aligned} \tag{3.2.41}$$

其中$\hat{N}_k=\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{N}/\sqrt{\lambda_k}$。由于$\hat{\boldsymbol{N}}$是 $\boldsymbol{N}$ 的线性变换，因此它同样是高斯随机向量，且满足

$$\begin{aligned}E\{\hat{N}_k\,\hat{N}_l\}&=E\{\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{N}\boldsymbol{v}_l^{\mathrm{T}}\boldsymbol{N}\}/\sqrt{\lambda_k\lambda_l}\\&=E\{\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{N}\boldsymbol{N}^{\mathrm{T}}\boldsymbol{v}_l\}/\sqrt{\lambda_k\lambda_l}\\&=\boldsymbol{v}_k^{\mathrm{T}}E\{\boldsymbol{N}\boldsymbol{N}^{\mathrm{T}}\}\boldsymbol{v}_l/\sqrt{\lambda_k\lambda_l}\\&=\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{\Sigma}_N\boldsymbol{v}_l/\sqrt{\lambda_k\lambda_l}=\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{v}_l\sqrt{\lambda_k/\lambda_l}\end{aligned} \tag{3.2.42}$$

上式利用了 $\boldsymbol{v}_l$ 是 $\boldsymbol{\Sigma}_N$ 的特征向量这一性质。根据 $\boldsymbol{v}_1$，…，$\boldsymbol{v}_n$ 的正交性，式(3.2.42)可以表示为

$$E\{\hat{N}_k\,\hat{N}_l\}=\begin{cases}1, & k=l\\0, & k\neq l\end{cases} \tag{3.2.43}$$

因此，$\hat{N}_1$，…，$\hat{N}_n$ 为服从$\mathcal{N}(0,1)$的独立同分布随机变量。对观测 $\boldsymbol{Y}$ 进行适当的线性变换后，可将非独立高斯噪声下的检测问题转化为独立同分布高斯噪声下的等价问题。式(3.2.39)给出了转换后的最优检测量。

本质上讲，上面所做计算的就是将最初在 $\mathbf{R}^n$ 上的标准坐标系转换成一个坐标轴平行于向量 $\boldsymbol{v}_1$，…，$\boldsymbol{v}_n$ 的坐标系。在变换后的坐标系中，噪声的坐标$\hat{N}_1$，…，$\hat{N}_n$ 是独立的。也可以从另一个角度来看待坐标系转换问题：令 $\boldsymbol{\Sigma}_N=\boldsymbol{B}^2$，其中 $\boldsymbol{B}=\sum_{k=1}^{n}\lambda^{1/2}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}$($\boldsymbol{\Sigma}_N$ 的平方根)。矩阵 $\boldsymbol{B}$ 的逆为 $\boldsymbol{B}^{-1}=\sum_{k=1}^{n}\lambda^{1/2}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}$ 且有 $\boldsymbol{\Sigma}_N^{-1}=\boldsymbol{B}^{-2}\triangleq(\boldsymbol{B}^{-1})^2$。如果定义 $\boldsymbol{s}_j^*=\boldsymbol{B}^{-1}\boldsymbol{s}_j$ 且 $y^*=\boldsymbol{B}^{-1}\boldsymbol{y}$，则有

$$(\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{y}=(\boldsymbol{s}_1^*-\boldsymbol{s}_0^*)^{\mathrm{T}}\boldsymbol{y}^* \tag{3.2.44}$$

此外，在 H_j 条件下，有 $\boldsymbol{Y}^*=\boldsymbol{N}^*+\boldsymbol{s}_j^*$，其中 $\boldsymbol{N}^*=\boldsymbol{B}^{-1}\boldsymbol{N}$。进而有

$$E\{\boldsymbol{N}^*(\boldsymbol{N}^*)^{\mathrm{T}}\}=E\{\boldsymbol{B}^{-1}\boldsymbol{N}\boldsymbol{N}^{\mathrm{T}}\boldsymbol{B}^{-1}\}=\boldsymbol{B}^{-1}E\{\boldsymbol{N}\boldsymbol{N}^{\mathrm{T}}\}\boldsymbol{B}^{-1}=\boldsymbol{B}^{-1}\boldsymbol{\Sigma}_N\boldsymbol{B}^{-1}=\boldsymbol{B}^{-1}\boldsymbol{B}\boldsymbol{B}\boldsymbol{B}^{-1}=\boldsymbol{I}$$

即 N_1^*，…，N_n^* 是服从$\mathcal{N}(0,1)$的独立同分布变量，它与上文中观测数据$\hat{\boldsymbol{Y}}$类似。事实上，

由于$\hat{\boldsymbol{Y}}=\sum_{k=1}^{n}\hat{Y}_k\boldsymbol{v}_k$ 和$\boldsymbol{Y}^*=\sum_{k=1}^{n}\hat{Y}_k\boldsymbol{e}_k$，因此$\boldsymbol{Y}^*$ 和$\hat{\boldsymbol{Y}}$是不同坐标系下的相同随机向量，其中$\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n$ 是$\mathbf{R}^n$ 空间的标准基向量，即$\boldsymbol{e}_k$ 是第k 个元素为1、其他元素为0的向量。

在独立同分布噪声条件下，以其他方式对观测向量$\boldsymbol{Y}$进行转换，也同样可以得到等效的观测数据。特别地，因为$\boldsymbol{\Sigma}_N$ 是正定的，也可以写作

$$\boldsymbol{\Sigma}_N=\boldsymbol{C}\boldsymbol{C}^{\mathrm{T}} \tag{3.2.45}$$

其中$\boldsymbol{C}$为一个$n\times n$的可逆下三角矩阵($\boldsymbol{C}$的所有上对角元素为0)。式(3.2.45)为$\boldsymbol{\Sigma}_N$ 的Cholesky 分解，有多种标准算法可从$\boldsymbol{\Sigma}_N$ 中求出$\boldsymbol{C}$[参考 Bierman(1977)]，且有$\boldsymbol{\Sigma}_N^{-1}=(\boldsymbol{C}^{\mathrm{T}})^{-1}\boldsymbol{C}^{-1}=(\boldsymbol{C}^{-1})^{\mathrm{T}}\boldsymbol{C}^{-1}$。在定义新的观测数据$\overline{\boldsymbol{Y}}=\boldsymbol{C}^{-1}\boldsymbol{Y}=\boldsymbol{C}^{-1}\boldsymbol{N}+\boldsymbol{C}^{-1}\boldsymbol{s}_j\triangleq\overline{\boldsymbol{N}}+\overline{\boldsymbol{s}}_j$ 后，可直接得到$\overline{\boldsymbol{N}}\sim\mathcal{N}(0,\boldsymbol{I})$。显然，我们再次得到独立同分布噪声情况下的最优检测器为$(\overline{\boldsymbol{s}}_1-\overline{\boldsymbol{s}}_0)^{\mathrm{T}}\overline{\boldsymbol{Y}}$。

值得注意的是，在上述转换过程中，$\boldsymbol{C}$是下三角矩阵，意味着$\boldsymbol{C}^{-1}$也是一个下三角矩阵。因此可以得到

$$\boldsymbol{y}_k=\sum_{l=1}^{k}h_{k,l}y_l \tag{3.2.46}$$

其中$h_{k,l}$是$\boldsymbol{C}^{-1}$的第$k-l$个元素。注意，式(3.2.46)对应因果运算，这意味着$\overline{y}_1$，…，$\overline{y}_n$可由y_1，…，y_n因果、线性滤波得到，且该滤波器可能是时变的。又因为该滤波器的输出噪声是白噪声(独立同分布)，所以也称为白化滤波器。式(3.2.24)对应的最优检测器结构可以表示为一个因果线性滤波器与一个相关器的级联(如图3.2.8所示)。其中，滤波器的脉冲响应为$\{h_{k,l}\}$，输入为y_1，…，y_n。相关器的作用是将观测数据y_1，…，y_n通过滤波器的输出与差分信号$(s_{11}-s_{01})$，…，$(s_{1n}-s_{0n})$通过相同结构滤波器的输出$(\overline{s}_{11}-\overline{s}_{01})$，…，$(\overline{s}_{1n}-\overline{s}_{0n})$作相关。

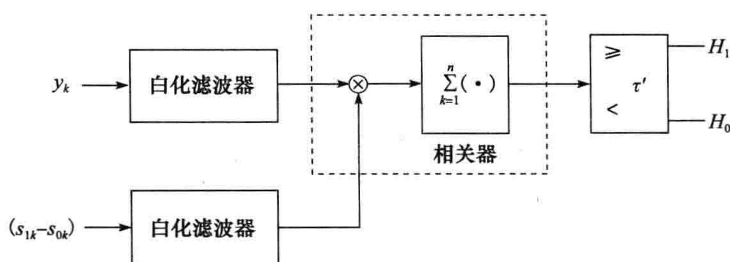


图3.2.8 非独立高斯噪声下相关信号的最优检测器

最后需要注意，信噪比$d^2=(\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{s}_1-\boldsymbol{s}_0)$可以用任意一个变换后的信号对进行表示，即

$$d^2=\|\hat{\boldsymbol{s}}_1-\hat{\boldsymbol{s}}_0\|^2=\|\boldsymbol{s}_1^*-\boldsymbol{s}_0^*\|^2=\|\overline{\boldsymbol{s}}_1-\overline{\boldsymbol{s}}_0\|^2 \tag{3.2.47}$$

由此可以发现，在非独立噪声下相干信号检测的性能依赖于信号在转换后的新坐标系(噪声分量满足独立同分布特性)下的距离[与式(3.2.37)比较]。注意，在式(3.2.47)中的三个信号对是等距的，这是因为它们都是相同向量对在不同坐标系下的表示，而坐标系之间仅存在旋转变换。

备注 3.2.3　信号选择

高斯噪声中的最优相干检测器可以通过增加$d^2=(\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}(\boldsymbol{s}_1-\boldsymbol{s}_0)$的值来提高性

能。在很多相干信号检测的具体应用中，通常来说在选择信号 $\boldsymbol{s}_0$ 和 $\boldsymbol{s}_1$ 时都会比较灵活。在这种情况下，选择能够最大化 d^2 的信号形式将更加合理。

正如之前的讨论中提到的，$\boldsymbol{\Sigma}_N^{-1}=\sum_{k=1}^{n}\lambda_k^{-1}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}$，其中 $\lambda_1,\cdots,\lambda_n$ 和 $\boldsymbol{v}_1,\cdots,\boldsymbol{v}_n$ 是特征值和对应的正交特征向量。因此，对于任意向量 $x\in\mathbf{R}^n$，都可以得到

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{x}=\sum_{k=1}^{n}\lambda_k^{-1}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{x}\leqslant\lambda_{\min}^{-1}\sum_{k=1}^{n}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{x}\tag{3.2.48}$$

其中 $\lambda_{\min}=\min(\lambda_1,\ \cdots,\ \lambda_n)$。由于

$$\sum_{k=1}^{n}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{x}^{\mathrm{T}}\left(\sum_{k=1}^{n}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{x}\right)=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=\|\boldsymbol{x}\|^2$$

可得

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Sigma}_N^{-1}\boldsymbol{x}\leqslant\lambda_{\min}^{-1}\|\boldsymbol{x}\|^2\tag{3.2.49}$$

需要注意的是，当且仅当 $\boldsymbol{x}$ 正比于最小特征值 $\lambda_{\min}$ 对应的特征向量时，式(3.2.49)才取等号。[如果 $\lambda_{\min}$ 对应多个特征向量，$\boldsymbol{x}$ 可以表示成这些特征向量的任意线性组合并且同样能够使式(3.2.49)成立，因为这些特征向量的线性组合仍是 $\lambda_{\min}$ 对应的特征向量。]

从以上分析中可以看出，对于一个固定的 $\|\boldsymbol{s}_1-\boldsymbol{s}_0\|$，选择差分信号 $\boldsymbol{s}_1-\boldsymbol{s}_0$ 的最优方式是沿着最小特征值对应的特征向量方向。$\boldsymbol{\Sigma}_N$ 中的特征值表示其对应特征向量方向上的噪声能量。因此，将信号沿着最小特征值对应特征向量的方向差分与沿着噪声最小的方向差分是等效的。通过这样处理后，可以得到 d^2 的值如下

$$d^2=\frac{1}{\lambda_{\min}}\|\boldsymbol{s}_1-\boldsymbol{s}_0\|^2\tag{3.2.50}$$

一旦选定了信号差分 $\boldsymbol{s}_1-\boldsymbol{s}_0$ 的方向，就可以进一步通过最大化 $\|\boldsymbol{s}_1-\boldsymbol{s}_0\|^2$ 来优化性能。显然，如果不对信号进行约束，$\|\boldsymbol{s}_1-\boldsymbol{s}_0\|^2$ 的值可以任意大。考虑到信号的总能量通常是固定的，在对式(3.2.50)的优化问题中需要加入这一约束。特别地，如果假设 $\|\boldsymbol{s}_1\|^2\leqslant P$ 和 $\|\boldsymbol{s}_2\|^2\leqslant P$，其中 $0<P<\infty$，可以得到

$$d^2=\frac{1}{\lambda_{\min}}(\boldsymbol{s}_1-\boldsymbol{s}_0)^{\mathrm{T}}(\boldsymbol{s}_1-\boldsymbol{s}_0)=\frac{1}{\lambda_{\min}}(\|\boldsymbol{s}_1\|^2-2\boldsymbol{s}_1^{\mathrm{T}}\boldsymbol{s}_0+\|\boldsymbol{s}_0\|^2)\tag{3.2.51}$$

其中，$\boldsymbol{s}_1^{\mathrm{T}}\boldsymbol{s}_0$ 是 $\boldsymbol{s}_0$ 和 $\boldsymbol{s}_1$ 的内积(点乘)。对于固定的 $\|\boldsymbol{s}_1\|^2$ 和 $\|\boldsymbol{s}_0\|^2$，如果 $\|\boldsymbol{s}_1\|$ 和 $\|\boldsymbol{s}_0\|$ 方向相反，则 $\boldsymbol{s}_1^{\mathrm{T}}\boldsymbol{s}_0$ 可以最小化，从而使得 d^2 最大化。例如 $\boldsymbol{s}_0=\alpha\boldsymbol{s}_1$，其中 $\alpha<0$。此时，可以得到

$$d^2=\frac{1}{\lambda_{\min}}(\|\boldsymbol{s}_1\|^2-2\alpha\|\boldsymbol{s}_1\|^2+\alpha^2\|\boldsymbol{s}_1\|^2)=\frac{\|\boldsymbol{s}_1\|^2(1-\alpha)^2}{\lambda_{\min}}\tag{3.2.52}$$

上式中 $\alpha=-\frac{\|\boldsymbol{s}_0\|}{\|\boldsymbol{s}_1\|}$。所以，对于固定的 $\|\boldsymbol{s}_0\|$ 和 $\|\boldsymbol{s}_1\|$，d^2 的最大值为

$$d^2=\frac{(\|\boldsymbol{s}_1\|+\|\boldsymbol{s}_0\|)^2}{\lambda_{\min}}$$

观察可知，通过选择 $\|\boldsymbol{s}_1\|^2=\|\boldsymbol{s}_0\|^2=P$，$d^2$ 的值可以被进一步最大化，此时有 $\alpha=-1$ 和

$$\max_{\|\boldsymbol{s}_j\|^2\leqslant P}d^2=\frac{4P}{\lambda_{\min}}\tag{3.2.53}$$

由于在选择 $\boldsymbol{s}_1-\boldsymbol{s}_0$ 时是沿着最小特征值对应的特征向量 $\boldsymbol{v}_{\min}$ 方向，因此可以通过选择 $\boldsymbol{s}_1=c\boldsymbol{v}_{\min}$ 和 $\boldsymbol{s}_0=-\boldsymbol{s}_1$ 来满足式(3.2.53)的要求，其中 c 的选取应保证 $\|\boldsymbol{s}_1\|^2=\|\boldsymbol{s}_0\|^2=P$，即 $c=$

$\sqrt{P}/\|\boldsymbol{v}_{\min}\|$，相应的最优信号由下式给出

$$\boldsymbol{s}_1 = \sqrt{P}\boldsymbol{v}_{\min}/\|\boldsymbol{v}_{\min}\|, \quad \boldsymbol{s}_0 = -\boldsymbol{s}_1 \tag{3.2.54}$$

下面举例说明上述概念。

例 3.2.4 双样本检测下的最优信号

在 $n=2$ 时，有

$$\boldsymbol{\Sigma}_N = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \tag{3.2.55}$$

其中 $|\rho|<1$。$\boldsymbol{\Sigma}_N$ 的特征值和对应的正交特征向量如下

$$\lambda_1 = \sigma^2(1-\rho), \quad \lambda_2 = \sigma^2(1+\rho)$$

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \boldsymbol{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{3.2.56}$$

因此，如果 $\rho>0$，则有 $\lambda_{\min}=\lambda_1$ 且最优信号为

$$\boldsymbol{s}_1 = \sqrt{P/2}\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \boldsymbol{s}_0 = \sqrt{P/2}\begin{pmatrix} -1 \\ 1 \end{pmatrix} \tag{3.2.57}$$

相应地，如果 $\rho<0$，则有 $\lambda_{\min}=\lambda_2$ 且最优信号为

$$\boldsymbol{s}_1 = \sqrt{P/2}\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \boldsymbol{s}_0 = \sqrt{P/2}\begin{pmatrix} -1 \\ -1 \end{pmatrix} \tag{3.2.58}$$

在这两种情况中，d^2 的最大值均为

$$d^2 = \frac{4P}{\sigma^2(1-|\rho|)} \tag{3.2.59}$$

信号的最优性可以由图(3.2.9)中的相同密度等高线得到。需要注意的是，这两种情况中信号向量都是在噪声密度下降最快的方向，因此对于固定的 P、σ^2 和 ρ，可以得到最大的信噪比。此外，就选择最优信号而言，只需要知道 ρ 的符号，而不需要其真实值。 ◀

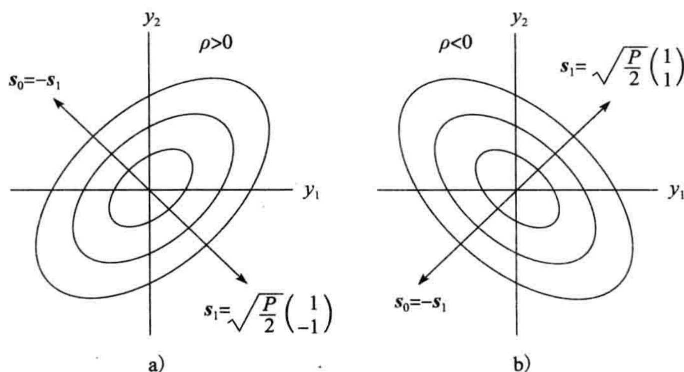


图 3.2.9 $\boldsymbol{\Sigma}_N=\sigma^2\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ 的高斯噪声中的最优信号

模型 3.2.3 随机参数信号检测

模型 3.2.1 和 3.2.2 讨论了两个信号对于观测者完全已知时的判定问题。然而在许多实际应用中，需要面临部分参数未知的信号判决问题。这类问题通常会出现在数字通信系

统中，例如在发射端将两个波形(分别表示“0”或“1”)之一调制到正弦载波上，而接收端则需要判定发射的是两者中的哪一个。虽然两个发信号的波形和载波频率在接收端是已知的，但是其幅度和载波相位却有可能是未知的。因此，在检测过程中必须考虑这些表征未知参数的量。类似地，在雷达、声呐以及其他一些应用中，需要检测信号的频率、到达时间，以及幅度等未知参数。

对于这一情况，为了方便起见，可以将式(3.2.1)写为

$$\begin{aligned} H_0 : Y_k &= N_k + s_{0k}(\theta), \quad k = 1,\cdots,n \\ H_1 : Y_k &= N_k + s_{1k}(\theta), \quad k = 1,\cdots,n \end{aligned} \tag{3.2.60}$$

其中，$\boldsymbol{s}_0(\theta)$和$\boldsymbol{s}_1(\theta)$是θ的已知向量值函数，其中θ是在参数空间Λ上取值的未知参数。假设θ随机(记作Θ)且在H_j假设下的密度为w_j，则式(3.2.60)的似然比可以写为

$$L(\boldsymbol{y}) = \frac{E_1\{p_N(\boldsymbol{y} - \boldsymbol{s}_1(\Theta))\}}{E_1\{p_N(\boldsymbol{y} - \boldsymbol{s}_0(\Theta))\}} = \frac{\int_\Lambda p_N(\boldsymbol{y} - \boldsymbol{s}_1(\Theta))\omega_1(\theta)\mu(\mathrm{d}\theta)}{\int_\Lambda p_N(\boldsymbol{y} - \boldsymbol{s}_0(\Theta))\omega_0(\theta)\mu(\mathrm{d}\theta)} \tag{3.2.61}$$

为了进一步分析，假设$\boldsymbol{s}_0(\theta)\equiv 0$，$\boldsymbol{s}_1(\theta)\triangleq \boldsymbol{s}(\theta)$(由于其他情况可进行类似处理，这里不作分析)。相应地，有

$$L(\boldsymbol{y}) = \int_\Lambda \frac{p_N(\boldsymbol{y} - \boldsymbol{s}(\theta))}{p_N(\boldsymbol{y})}\omega(\theta)\mu(\mathrm{d}\theta) = \int_\Lambda L_\theta(\boldsymbol{y})\omega(\theta)\mu(\mathrm{d}\theta) \tag{3.2.62}$$

其中$L_\theta(\boldsymbol{y})$是在$\Theta=\theta$条件下的似然比，并且去掉了w_1的下标。从式(3.2.62)可以发现：此时的$L(\boldsymbol{y})$只是对于已知θ的简单平均似然比。在θ已知时，式(3.2.60)是一个确定性的信号问题，因此$L_\theta(\boldsymbol{y})$可以利用模型 3.2.1 和 3.2.2 的方法直接求解。例如，对于服从$\mathcal{N}(0, \sigma^2)$的独立同分布噪声样本，$L(\boldsymbol{y})$变为

$$L(\boldsymbol{y}) = \int_\Lambda \exp\left\{\left[\boldsymbol{s}^{\mathrm{T}}(\theta)\boldsymbol{y} - \frac{1}{2}\|\boldsymbol{s}(\theta)\|^2\right]/\sigma^2\right\}\omega(\theta)\mu(\mathrm{d}\theta) \tag{3.2.63}$$

类似地，对于非独立同分布高斯噪声情况，式(3.2.63)同样成立。其中$\sigma^2=1$，$\boldsymbol{y}$、$\boldsymbol{s}(\theta)$是经过预白化后的变量。一些应用中可能会出现这类问题，具体内容将在下面的例子中进一步说明。

例 3.2.5　正弦载波调制信号的非相干检测

考虑信号对$\boldsymbol{s}_0(\theta)=0$和$\boldsymbol{s}_1(\theta)=\boldsymbol{s}(\theta)$，设

$$s_k(\theta) = \alpha_k \sin[(k-1)\omega_c T_s + \theta], \quad k = 1,\cdots,n \tag{3.2.64}$$

其中，a_1，a_2，…，a_n为信号的幅度序列，Θ是与噪声独立且均匀分布在$[0, 2\pi]$区间上的随机相角，ω_c和T_s分别表示载波频率和采样间隔，满足$n\omega_c T_s = m2\pi$且m为整数(在时间段$[0, nT_s]$内有整数个正弦波周期)。再假设正弦波每个周期的采样数n/m是大于 1 的整数。下面以开关键控(OOK)为例来说明具体的信号产生模型。在这种情况下，$[0, nT_s]$间隔内什么都不发射代表“0”，而发射$a(t)$调制的正弦波代表“1”，此时，序列a_1，a_2，…，a_n是波形$a(t)$的采样，θ表示接收端未知的载波相角。接收端载波相位未知的调制载波检测称为非相干检测。通常来说，假设相角在$[0, 2\pi]$区间上服从均匀分布，即所有的相角都有可能发生，如果没有与本假设相矛盾的先验信息，假设也是合理的。

假设噪声独立同分布并且服从$\mathcal{N}(0, \sigma^2)$分布，利用式(3.2.63)，检测问题的似然比

可以写为

$$L(\mathbf{y}) = \frac{1}{2\pi}\int_0^{2\pi} \exp\left\{\frac{1}{\sigma^2}\left(\sum_{k=1}^{n} y_k s_k(\theta) - \frac{1}{2}\sum_{k=1}^{n} s_k^2(\theta)\right)\right\}\mathrm{d}\theta \tag{3.2.65}$$

利用 $\sin(a+b)=\cos a\sin b+\sin a\cos b$ 这一性质，式(3.2.65)指数项括号中的第一项可以写为 $\sum_{k=1}^{n} y_k s_k(\theta) = y_c\sin\theta + y_s\cos\theta$,其中

$$y_c \triangleq \sum_{k=1}^{n} a_k y_k \cos[(k-1)\omega_c T_s], \quad y_s \triangleq \sum_{k=1}^{n} a_k y_k \sin[(k-1)\omega_c T_s] \tag{3.2.66}$$

类似地，利用 $\sin^2 a=\frac{1}{2}-\frac{1}{2}\cos 2a$ 这一性质，指数项括号中的第二项可以写为

$$-\frac{1}{2}\sum_{k=1}^{n} s_k^2(\theta) = -\frac{1}{4}\sum_{k=1}^{n} a_k^2 + \frac{1}{4}\sum_{k=1}^{n} a_k^2\cos(2(k-1)\omega_c T_s + 2\theta) \tag{3.2.67}$$

在大多数实际情况下，式(3.2.67)右边第二项对于所有 θ 均为 0 或近似为 0。例如，如果信号序列 a_1，a_2，…，a_n 是常数乘以±1 序列，或者 a_1，a_2，…，a_n 具有升余弦形式 $a_k=A[1-\cos((k-1)2\pi/(n-1))]$，$k=1$，…，$n$，那么第二项就为 0。在其他情况下，相对于两倍杂波频率而言，a_1^2，…，a_n^2 变换缓慢。所以第二项就等同于对高频信号进行低通滤波，其输出是可以忽略的。在任何情况下，均假设 a_1，a_2，…，a_n 使得第二项对所 θ 均为 0，那么 $L(\mathbf{y})$将变为

$$L(\mathbf{y}) = \mathrm{e}^{-n\overline{a^2}/4\sigma^2} \times \frac{1}{2\pi}\int_0^{2\pi} \exp\left\{\frac{1}{\sigma^2}(y_c\sin\theta + y_s\cos\theta)\right\}\mathrm{d}\theta \tag{3.2.68}$$

其中 $\overline{a^2} = \frac{1}{n}\sum_{k=1}^{n} a_k^2$。

式(3.2.68)与例 2.5.1 中的似然比相似，比较式(3.2.68)和式(2.5.17)可以得到

$$L(\mathbf{y}) = \mathrm{e}^{-n\overline{a^2}/4\sigma^2} I_0(r/\sigma^2) \tag{3.2.69}$$

其中 $r=[y_c^2+y_s^2]^{1/2}$ 和 I_0 是第一类零阶修正贝塞尔函数。考虑到 $I_0(\cdot)$具有单调性，最优检测器可以写为

$$\tilde{\delta}_o(\mathbf{y}) = \begin{cases} 1, & r > \tau' \triangleq \sigma^2 I_0^{-1}(\tau \mathrm{e}^{n\overline{a^2}/4\sigma^2}) \\ \gamma, & r = \tau' \triangleq \sigma^2 I_0^{-1}(\tau \mathrm{e}^{n\overline{a^2}/4\sigma^2}) \\ 0, & r < \tau' \triangleq \sigma^2 I_0^{-1}(\tau \mathrm{e}^{n\overline{a^2}/4\sigma^2}) \end{cases} \tag{3.2.70}$$

该检测器的结构如图 3.2.10 所示。需要注意的是，观测信号 y_1，…，y_n 被分离到两个通道：第一个通道对每个 y_k 乘以 $\cos[(k-1)\omega_c T_s]$；另一个通道对每个 y_k 乘以 $\sin[(k-1)\omega_c T_s]$（这两个通道分别称为同相和正交通道）。每个通道将乘积的结果与幅度序列 a_1，a_2，…，a_n 进行相关后，对相关的结果再进行组合得到 r，并与阈值进行比较（该结构通常也称为包络检测器）。直观上看，对于一个给定信号，检测器的每个通道都会获得一定的信号能量，具体值取决于载波的实际相位。但是，无论载波相位如何变化，联合两个信道后便能获得信号的全部能量。

为了分析式(3.2.70)检测器的性能，需要计算 $P_j(R>\tau')=P_j(R^2>(\tau')^2)$，其中 $j=0$，1。由于 $R^2=Y_c^2+Y_s^2$，其中，$Y_c \triangleq \sum_{k=1}^{n} a_k Y_k\cos[(k-1)\omega_c T_S]$，$Y_s \triangleq \sum_{k=1}^{n} a_k Y_k\sin[(k-1)\omega_c T_S]$，所需

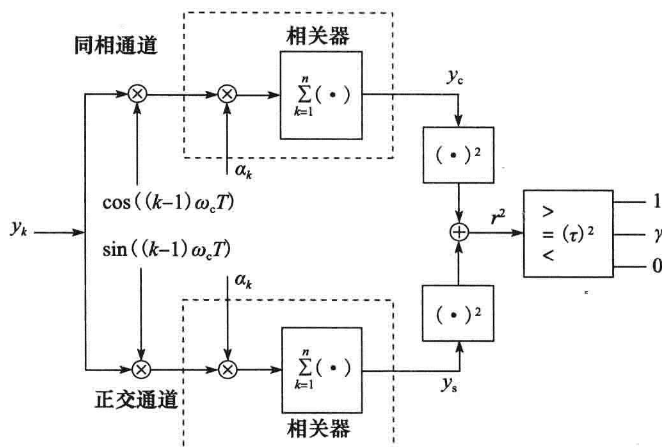


图 3.2.10　高斯噪声中正弦波调制信号的非相干最优检测器

要求解的概率可以从 Y_c 和 Y_s 在两种假设下的联合概率密度函数得到。在 H_0 假设下，$\boldsymbol{Y}$ 服从$\mathcal{N}(0,\sigma^2\boldsymbol{I})$分布，由于 Y_c 和 Y_s 相对于 $\boldsymbol{Y}$ 均是线性函关系，从而它们在 H_0 下是联合高斯的。因此，可以通过 Y_c 和 Y_s 的均值和方差以及它们之间的相关系数来确定(Y_c,Y_s)在 H_0 下的的联合密度。经过计算，均值可表示为

$$E\{Y_c\,|\,H_0\}=\sum_{k=1}^{n}a_k E\{N_k\}\cos[(k-1)\omega_c T_s]=0$$

方差为

$$\begin{aligned}\operatorname{var}[Y_c\,|\,H_0]=E\{Y_c^2\,|\,H_0\}&=\sum_{k=1}^{n}\sum_{l=1}^{n}a_k a_l E\{N_k N_l\}\times\cos[(k-1)\omega_c T_s]\cos[(l-1)\omega_c T_s]\\&=\sigma^2\sum_{k=1}^{n}a_k^2\cos^2[(k-1)\omega_c T_s]=\frac{n\sigma^2\,\overline{a^2}}{2}\end{aligned}$$

其中再次利用了式(3.2.67)中第二项对于所有 θ 均为零这一假设。类似地，有 $E\{Y_s\,|\,H_0\}=0$ 和 $\operatorname{var}\{Y_s\,|\,H_0\}=\operatorname{var}\{Y_c\,|\,H_0\}$。$Y_c$ 和 Y_s 在 H_0 假设下的相关系数为 $\operatorname{cov}(Y_c,Y_s\,|\,H_0)/[\operatorname{var}(Y_c\,|\,H_0)\operatorname{var}(Y_s\,|\,H_0)]^{1/2}$。由于 Y_c 和 Y_s 在 H_0 下均值为 0，因此有

$$\begin{aligned}\operatorname{cov}(Y_c,Y_s\,|\,H_0)&=E\{Y_cY_s\,|\,H_0\}\\&=\sum_{k=1}^{n}\sum_{l=1}^{n}a_k a_l E\{N_k N_l\}\cos[(k-1)\omega_c T_s]\sin[(l-1)\omega_c T_s]\\&=\sigma^2\sum_{k=1}^{n}a_k^2\cos[(k-1)\omega_c T_s]\sin[(k-1)\omega_c T_s]\\&=\frac{\sigma^2}{2}\sum_{k=1}^{n}a_k^2\sin[2(k-1)\omega_c T_s]=0\end{aligned}\tag{3.2.71}$$

由式(3.2.71)易知 H_0 假设下 Y_c 和 Y_s 是不相关的。此外，由于它们是联合高斯的，从而也是独立的。

综上所述，在 H_0 下，Y_c 和 Y_s 是独立且服从$\mathcal{N}(0,n\sigma^2\,\overline{a^2}/2)$分布的随机变量。需要注意的是，它们的随机化是不相关的，所以式(3.2.70)的虚警概率变为

$$P_0(\Gamma_1)=\iint_{\{y_c^2+y_s^2\geqslant(\tau')^2\}}\frac{1}{\pi n\sigma^2\overline{a^2}}\mathrm{e}^{-(y_c^2+y_s^2)/n\sigma^2\overline{a^2}}\mathrm{d}y_c\mathrm{d}y_s$$
$$=\frac{1}{\pi n\sigma^2\overline{a^2}}\int_0^{2\pi}\int_{\tau'}^{\infty}r\mathrm{e}^{-r^2/n\sigma^2\overline{a^2}}\mathrm{d}r\mathrm{d}\psi$$
$$=\mathrm{e}^{-(\tau')^2/n\sigma^2\overline{a^2}} \tag{3.2.72}$$

其中，第二个等号引入了极坐标系下的 r 和 ψ。

为了确定式(3.2.70)的检测概率，需要获得 Y_c 和 Y_s 在 H_1 下的联合密度。对于给定的 $\Theta=\theta$，$\boldsymbol{Y}$ 在 H_1 下服从$\mathcal{N}(\boldsymbol{s}(\theta),\sigma^2\boldsymbol{I})$的条件分布。因此，对于给定的 $\Theta=\theta$，Y_c 和 Y_s 是条件联合高斯的，进而可以直接得到

$$E\{Y_c\,|\,H_1,\Theta=\theta\}=\sum_{k=1}^{n}a_kE\{Y_k\,|\,H_1,\Theta=\theta\}\cos[(k-1)\omega_cT_s]$$
$$=\sum_{k=1}^{n}a_k^2\sin[(k-1)\omega_cT_s+\theta]\cos[(k-1)\omega_cT_s]$$
$$=\frac{n\overline{a^2}}{2}\sin\theta$$

类似地，还可以得到 $E\{Y_s\,|\,H_1,\Theta=\theta\}=(n\overline{a^2}/2)\cos\theta$。当 θ 固定时，由于 $\boldsymbol{Y}$ 仅改变了均值，因此 Y_c 和 Y_s 在 H_1 下的方差和协方差和 H_0 假设相同。

Y_c 和 Y_s 在 H_1 下的全概率密度可以通过对 θ 的条件密度进行平均得到，即

$$p_{Y_c,Y_s}(y_c,y_s\,|\,H_1)=\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{\pi n\sigma^2\overline{a^2}}\exp\left\{-\frac{1}{n\sigma^2\overline{a^2}}q(y_c,y_s;\frac{n\overline{a^2}}{2},\theta)\right\}\mathrm{d}\theta$$
$$=p_{Y_c,Y_s}(y_c,y_s\,|\,H_0)\mathrm{e}^{-n\overline{a^2}/4\sigma^2}I_0\left(\frac{[y_c^2+y_s^2]^{1/2}}{\sigma^2}\right) \tag{3.2.73}$$

其中，$p_{Y_c,Y_s}(y_c,\ y_s\,|\,H_0)$是 Y_c 和 Y_s 在 H_0 下的联合密度，q 与式(2.5.14)中定义的 q 相同。因此，式(3.2.70)的检测概率为

$$P_D(\hat{\delta}_o)=P_1(\Gamma_1)=\iint_{\{y_c^2+y_s^2\geqslant(\tau')^2\}}p_{Y_c,Y_s}(y_c,y_s\,|\,H_1)\mathrm{d}y_c\mathrm{d}y_s$$
$$=\frac{\mathrm{e}^{-n\overline{a^2}/4\sigma^2}}{\pi n\sigma^2\overline{a^2}}\int_0^{2\pi}\int_{\tau'}^{\infty}r\mathrm{e}^{-r^2/n\sigma^2\overline{a^2}}I_0\left(\frac{r}{\sigma^2}\right)\mathrm{d}r\mathrm{d}\psi$$
$$=\int_{\tau_o}^{\infty}x\mathrm{e}^{-(x^2+b^2)/2}I_0(bx)\mathrm{d}x\triangleq Q(b,\tau_o) \tag{3.2.74}$$

其中第三个等式定义了 $b^2=n\overline{a^2}/2\sigma^2$ 和 $\tau_o=\tau'/\sigma^2b$，并用 x 替换 τ/σ^2b。式(3.2.74)中定义的函数称为 Q 函数。此外，在式(3.2.72)中，$Q(0,\ \tau_o)=P_F(\tilde{\delta}_o)$。

由式(3.2.72)容易设定虚警概率为 α 的尼曼-皮尔逊检测器的判决阀值为

$$\tau'=[n\sigma^2\overline{a^2}\log(1/\alpha)]^{1/2} \tag{3.2.75}$$

从上式可以得到式(3.2.74)的 ROC 为

$$P_D(\hat{\delta}_o)=Q[b,[2\log(1/\alpha)]^{1/2}] \tag{3.2.76}$$

接收器的这些工作特性与相干问题(见图 2.4.4)非常相似，在 Helstrom(1986)中的图 5.4.3也可以找到其特性曲线。

从式(3.2.76)可以看出，尼曼-皮尔逊检测器的性能仅依赖于参数 b。考虑到平均信号

能量为

$$E\left\{\frac{1}{n}\sum_{k=1}^{n}s_k^2(\Theta)\right\}=\frac{1}{n}\frac{1}{2\pi}\int_0^{2\pi}\sum_{k=1}^{n}a_k^2\sin^2[(k-1)\omega_c T_s+\theta]\mathrm{d}\theta=\frac{\overline{a^2}}{2}$$

因此，与相干检测中的 d^2 相似，$b^2=\frac{n\overline{a^2}}{2\sigma^2}$也可以理解为信噪比。如果检测相同的相干信号（$\theta$已知），则 d^2 相应的取值为

$$d^2=\frac{1}{\sigma^2}\sum_{k=1}^{n}s_k^2(\theta)=\frac{n\overline{a^2}}{2\sigma^2}=b^2 \tag{3.2.77}$$

因此，这两个信噪比实际上是相同的。但是，对于固定的 α 来说，两个检测器的性能在 $b=d$ 时是不同的。考虑多数实际情况中 SNR 和 α 的范围后，有

$$Q[b,[2\log(1/\alpha)]^{1/2}]=1-\Phi[\Phi^{-1}(1-\alpha)-d] \tag{3.2.78}$$

其中 $b\approx d+0.4$[见 Helstrom(1968)]。因此，相对于相干检测技术来说，如果采用非相干检测技术，就需要更高的 SNR 来达到相同的性能。相干检测的缺点在于需要利用其他途径来得到载波相位。

总结上例可以发现，似然比推导过程可以用在信号形式与式(3.2.6)相同，但不同假设下的幅度序列存在差异的模型中。例如

$$s_{jk}(\theta)=a_{jk}\sin[(k-1)\omega_c T_s+\theta],\quad k=1,\cdots,n \tag{3.2.79}$$

其中 $j=0$，1。对原假设检验问题增加仅有噪声的第三种假设后，独立同分布噪声下的似然比可由式(3.2.69)得到。特别地

$$L(\boldsymbol{y})=\frac{E_1\{p_N(\boldsymbol{y}-\boldsymbol{s}_1(\Theta))\}}{E_0\{p_N(\boldsymbol{y}-\boldsymbol{s}_0(\Theta))\}}=\frac{E_1\{p_N(\boldsymbol{y}-\boldsymbol{s}_1(\Theta)/p_N(\boldsymbol{y}))\}}{E_0\{p_N(\boldsymbol{y}-\boldsymbol{s}_0(\Theta)/p_N(\boldsymbol{y}))\}}=\frac{L_1(\boldsymbol{y})}{L_0(\boldsymbol{y})}$$

其中 $L_j(\boldsymbol{y})$是第三种假设相对 H_j 的似然比。由式(3.2.69)可以得到

$$L(y)=\frac{\mathrm{e}^{-n\overline{a_1^2}/4\sigma^2}I_0(r_1/\sigma^2)}{\mathrm{e}^{-n\overline{a_0^2}/4\sigma^2}I_0(r_0/\sigma^2)} \tag{3.2.80}$$

其中，$\overline{a_j^2}=(1/n)\sum_{k=1}^{n}a_{jk}^2$，$r_j=[y_{cj}^2+y_{sj}^2]^{1/2}$，且有 $y_{cj}=\sum_{k=1}^{n}a_{jk}y_k\cos[(k-1)\omega_c T_s]$，$y_{sj}=\sum_{k=1}^{n}a_{jk}y_k\sin[(k-1)\omega_c T_s]$。

可以发现，最优检测器需要将两个图 3.2.10 所示的检测器的输出进行联合处理，每个检测器“匹配”一个信号幅度序列。例如，如果信号的能量相等($\overline{a_0^2}=\overline{a_1^2}$)并且假设均匀代价和相同的先验概率($\tau=1$)，由于 I_0 是单调递增的，相应的贝叶斯检验变为

$$\delta_{\mathrm{B}}(\boldsymbol{y})=\begin{cases}1, & r_1>r_0\\ 0\text{ 或 }1, & r_1=r_0\\ 0, & r_1<r_0\end{cases} \tag{3.2.81}$$

如果假设幅度序列正交，例如

$$\sum_{k=1}^{n}a_{jk}a_{lk}=0,\quad j\neq l$$

以及

$$\sum_{k=1}^{n} a_{jk} a_{lk} \sin[(k-1)\omega_c T_s + \theta] = 0$$

对于所有 θ 成立，则误差概率可以直接写为(见习题 11)

$$P_e = \frac{1}{2} e^{-b^2/4} \tag{3.2.82}$$

其中，$b^2 = n\overline{a_1^2}/2\sigma^2 (\equiv n\overline{a_0^2}/2\sigma^2)$。

上述检测问题的其他方面，包括幅度和相位未知时的检测将在习题中进一步研究。

模型 3.2.4　统计信号检测

在某些应用中，信号通常完全随机或需要进行统计建模。这类模型是非常有用的，例如在射电天文、声呐或者其他应用中，信号在传输过程中会受到传播介质扰动或多径等干扰。对于这一情况，通常使用式(3.2.1)中的模型和式(3.2.4)中的似然比函数。

此类问题的一个特例是信号和噪声均为高斯随机变量。为了研究这一情况，先给出在 $\mathbf{R}^n$ 上的假设检验问题如下

$$\begin{aligned} H_0: \boldsymbol{Y} &\sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \\ H_1: \boldsymbol{Y} &\sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \end{aligned} \tag{3.2.83}$$

对数似然比可以写为

$$\begin{aligned} \log L(\boldsymbol{y}) &= \frac{1}{2}\log[|\boldsymbol{\Sigma}_0|/|\boldsymbol{\Sigma}_1|] + \frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_0) \\ &\quad - \frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu}_1)^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_1) \\ &= \frac{1}{2}\boldsymbol{y}^{\mathrm{T}}[\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{\Sigma}_1^{-1}]\boldsymbol{y} + [\boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1} - \boldsymbol{\mu}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}]\boldsymbol{y} + C \end{aligned} \tag{3.2.84}$$

其中 $C = \frac{1}{2}(\log(|\boldsymbol{\Sigma}_0|/|\boldsymbol{\Sigma}_1|) + \boldsymbol{\mu}_0^{\mathrm{T}}\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1^{\mathrm{T}}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1)$。值得注意的是，$L(\boldsymbol{y})$ 由 $\boldsymbol{y}$ 的二次项、$\boldsymbol{y}$ 的一次项及常数项构成。如果这两个方差相同，即 $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}$，那么二次项就会消失；又因为 C 可以合并到阈值中，从而在本质上可得到一个线性统计检测 $(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}$。以上是在高斯噪声下的相干检测情况，与模型 3.2.2 相同。对于另一种情况，均值向量在两种假设检验下是相同的($\boldsymbol{\mu}_0 = \boldsymbol{\mu}_1$)，为了不失一般性，可以把它们设成 0，此时 $\log L(\boldsymbol{y})$ 变成二次型。

后一种情况适用于检测高斯噪声背景下的 0 均值随机信号。特别地，考虑到假设对

$$\begin{aligned} H_0: \boldsymbol{y} &= \boldsymbol{N} \\ H_1: \boldsymbol{y} &= \boldsymbol{N} + \boldsymbol{S} \end{aligned} \tag{3.2.85}$$

其中，$\boldsymbol{N} \sim \mathcal{N}(0, \sigma^2\boldsymbol{I})$，$\boldsymbol{S} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_s)$。噪声为非独立同分布的高斯分布时，预白化处理后的假设检验问题仍能用式(3.2.85)描述，这是因为高斯信号 $\boldsymbol{S}$ 的线性变换仍服从高斯分布。此外，每个假设都存在信号时，各个假设条件下的信号可在增加仅有噪声的假设后由式(3.2.85)得到。

如果假设信号和噪声总是独立的，那么式(3.2.85)中的假设对是式(3.2.83)在 $\boldsymbol{\Sigma}_0 = \sigma^2\boldsymbol{I}$ 和 $\boldsymbol{\Sigma}_1 = \sigma^2\boldsymbol{I} + \boldsymbol{\Sigma}_s$ 下的一个特例。因此，由式(3.2.84)可以得到式(3.2.85)的最优检测器具有如下形式

$$\widetilde{\delta}_{\text{o}}(\boldsymbol{y})=\begin{cases}1, & \boldsymbol{y}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{y}>\tau' \\ r, & \boldsymbol{y}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{y}=\tau' \\ 0, & \boldsymbol{y}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{y}<\tau'\end{cases} \tag{3.2.86}$$

其中，$\tau'\triangleq 2(\log\tau-C)$，$\boldsymbol{Q}\triangleq\sigma^{-2}\boldsymbol{I}-(\sigma^2\boldsymbol{I}+\boldsymbol{\Sigma}_{\mathrm{s}})^{-1}\equiv\sigma^{-2}\boldsymbol{\Sigma}_{\mathrm{s}}(\sigma^2\boldsymbol{I}+\boldsymbol{\Sigma}_{\mathrm{s}})^{-1}$。从式(3.2.86)中可以看到，最优检测器通过将二次型 $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{y}$ 的值与阈值进行比较来实现判决。该结构也称为二次型检测器。

例如，如果信号样本是服从$\mathcal{N}(0,\sigma_{\mathrm{s}}^2)$的独立同分布随机变量，则有 $\boldsymbol{\Sigma}_{\mathrm{s}}=\sigma_{\mathrm{s}}^2\boldsymbol{I}$ 和

$$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{y}=\frac{\sigma_{\mathrm{s}}^2}{\sigma^2(\sigma^2+\sigma_{\mathrm{s}}^2)}\sum_{k=1}^{n}y_k^2 \tag{3.2.87}$$

因此，在这一特殊情况下，最优检测器其实是将 $\sum_{k=1}^{n}y_k^2$ 的值和阈值进行比较。由于$\frac{1}{n}\sum_{k=1}^{n}y_k^2$是观测波形的平均能量，因此该检测器结构有时也称为能量检测器(也称辐射计)。这种检测方式是非常合理的，因为在两种假设下，随机观测数据仅在平均能量上存在差异。该问题可以和白噪声背景中判决两个常数信号 $\boldsymbol{\mu}_0=s_0\mathbf{1}$ 和 $\boldsymbol{\mu}_1=s_1\mathbf{1}$ 形成对比，其中，$\mathbf{1}=(1,\cdots,1)^{\mathrm{T}}$，$s_1>s_0$。此时，最优检测器将$\sum_{k=1}^{n}y_k$ 和阈值进行比较。需要注意的是，由于两种假设下随机观测数据的不同之处在于其(统计意义上）平均幅度存在差别，因此通过将平均观测数据$\frac{1}{n}\sum_{k=1}^{n}y_k$ 和阈值进行比较来实现判决是合理的。该检测结构通常也称为线性检测器。

为了分析式(3.2.86)中检测器的性能，需要分别计算 $j=0$，1 时的概率 $P_j(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{Y}>\tau')$($j=0$对应虚警概率，$j=1$ 对应检测概率)。如果首先对观测数据进行与相关检测类似的变换处理，那么求解虚警概率和检测概率将会更容易。特别地，假设 λ_1，…，λ_n 和 $\boldsymbol{v}_1$，…，$\boldsymbol{v}_n$ 是信号协方差矩阵 $\boldsymbol{\Sigma}_{\mathrm{s}}$ 的特征值和对应的正交特征向量。则有 $\boldsymbol{\Sigma}_{\mathrm{s}}=\sum_{k=1}^{n}\lambda_k\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}$，并且可以写作

$$\boldsymbol{I}=\sum_{k=1}^{n}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}$$

从而得到

$$(\sigma^2\boldsymbol{I}+\boldsymbol{\Sigma}_{\mathrm{s}})^{-1}=\sum_{k=1}^{n}(\sigma^2+\lambda_k)^{-1}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}$$

此外，还可以得到

$$\boldsymbol{Q}=\sum_{k=1}^{n}\frac{1}{\sigma^2}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}-\sum_{k=1}^{n}\frac{1}{\sigma^2+\lambda_k}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}}=\sum_{k=1}^{n}\frac{\lambda_k}{\sigma^2(\sigma^2+\lambda_k)}\boldsymbol{v}_k\boldsymbol{v}_k^{\mathrm{T}} \tag{3.2.88}$$

进而，统计检测可以写为

$$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{y}=\sum_{k=1}^{n}(\overline{y_k})^2 \tag{3.2.89}$$

其中 $\overline{\boldsymbol{y}}_k\triangleq[\lambda_k/\sigma^2(\sigma^2+\lambda_k)]^{1/2}\boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{y}$。

由于 $\boldsymbol{v}_1$，…，$\boldsymbol{v}_n$ 是对于$\sigma^2\boldsymbol{I}$ 和$(\sigma^2\boldsymbol{I}+\boldsymbol{\Sigma}_{\mathrm{s}})$的一组正交特征向量，与式(3.2.42)类似，从中可以得到以下结论：$\overline{Y}_1$，…，$\overline{Y}_n$ 在两种假设下均为独立的零均值高斯随机变量，其方差为

$$\sigma_{jk}^2 \triangleq \operatorname{var}(\overline{Y}_k \mid H_j) = \begin{cases} \dfrac{\lambda_k}{\sigma^2+\lambda_k}, & j=0 \\ \dfrac{\lambda_k}{\sigma^2}, & j=1 \end{cases} \tag{3.2.90}$$

这说明在 H_j 下，$\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{Y}$ 是独立随机变量的和，且求和中的第 k 项 Y_k^2 是服从$\mathcal{N}(0, \sigma_{jk}^2)$分布的随机变量的平方。$T_k \triangleq \overline{Y}_k^2$ 在 H_j 下的概率密度函数，可以表示为[参见 Papoulis (1986)]

$$p_{T_k}(t \mid H_j) = \begin{cases} \dfrac{1}{\sqrt{2\pi t}\sigma_{jk}} \mathrm{e}^{-t/2\sigma_{jk}^2}, & t>0 \\ 0, & t \leqslant 0 \end{cases} \tag{3.2.91}$$

式(3.2.91)为$\gamma\left(\dfrac{1}{2}, \dfrac{1}{2\sigma_{jk}^2}\right)$密度。$T \triangleq \sum_{k=1}^{n} \overline{y}_k^2$ 的概率密度 p_T 是n重卷积 $p_{T_1} * p_{T_2} * \cdots * p_{T_n}$，用傅里叶变换表示会更容易，即

$$p_T = f^{-1}\left\{\prod_{k=1}^{n} \phi_{T_k}\right\} \tag{3.2.92}$$

其中 $\phi_{T_k}(u) = F\{p_{T_k}\}(u) = E\{\mathrm{e}^{\mathrm{i}uT_k}\}(i=\sqrt{-1})$是 T_k 的特征函数。$\gamma(1/2, 1/2\sigma_{jk}^2)$的特征函数可以表示为[参见 Lukacs(1960)]

$$\phi_{T_k}(u) = \frac{1}{[1-2\mathrm{i}u\sigma_{jk}^2]^{1/2}}, \quad u \in \mathbf{R} \tag{3.2.93}$$

因此

$$p_T(t \mid H_j) = \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}ut} \prod_{k=1}^{n} [1-2\mathrm{i}u\sigma_{jk}^2]^{-1/2} \mathrm{d}u \tag{3.2.94}$$

式(3.2.94)没有闭式表达式。但是，在 $\sigma_{j1}^2 = \cdots = \sigma_{jn}^2 \triangleq \sigma_j^2$ 这一特殊情况下，式(3.2.94)能够转化为

$$p_T(t \mid H_j) = \begin{cases} \dfrac{1}{(2\sigma_j^2)^{n/2}\Gamma(n/2)} t^{(n/2-1)} \mathrm{e}^{-t/2\sigma_j^2}, & t>0 \\ 0, & t \leqslant 0 \end{cases} \tag{3.2.95}$$

其中，$\Gamma(x) = \int_0^{\infty} \mathrm{e}^{-y} y^{x-1} \mathrm{d}y$ 是伽马函数。这与 $\lambda_1 = \lambda_2 = \cdots = \lambda_n \triangleq \sigma_s^2$ 时的情况类似，进而可以得到

$$\boldsymbol{\Sigma}_s = \sigma_s^2 \sum_{k=1}^{n} \boldsymbol{v}_k \boldsymbol{v}_k^{\mathrm{T}} = \sigma_s^2 \boldsymbol{I}$$

又根据信号样本服从 $N(0, \sigma_s^2)$分布且具有独立同分布特性，因此有 $\sigma_0^2 = \sigma_s^2/(\sigma^2+\sigma_s^2)$和$\sigma_1^2 = \sigma_s^2/\sigma^2$。

式(3.2.95)是 $\Gamma\left(\dfrac{n}{2}, \dfrac{1}{2\sigma_{jk}^2}\right)$的密度，由此可以得到

$$P_j(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{Y} > \tau') = 1 - \Gamma\left(\frac{n}{2}; \frac{\tau'}{2\sigma_j^2}\right) \tag{3.2.96}$$

其中

$$\Gamma(x;t) \triangleq \int_0^t \mathrm{e}^{-y} y^{x-1} \mathrm{d}y / \Gamma(x)$$

这是一个不完全伽马函数[⊖]。对于虚警概率为 α 的尼曼-皮尔逊检测器而言，可以选择

$$\tau' = 2\sigma_0^2\Gamma^{-1}\left(\frac{n}{2};1-\alpha\right)$$

其中 $\Gamma^{-1}(x;\ \cdot)$ 是 $\Gamma(x;\ \cdot)$ 关于第二个变量的反函数。ROC 可以写作

$$P_{\mathrm{D}}(\widetilde{\delta}_{\mathrm{NP}}) = 1-\Gamma\left[\frac{n}{2};\frac{\sigma_0^2}{\sigma_1^2}\Gamma^{-1}\left(\frac{n}{2};1-\alpha\right)\right] \tag{3.2.97}$$

因此，检测性能可以由 n 和 $\sigma_0^2/\sigma_1^2=1/(1+\sigma_s^2/\sigma^2)$ 两个参数来衡量。需要注意的是：σ_s^2/σ^2 是平均信号能量与平均噪声能量之比，且式(3.2.84)的检测性能与 σ_s^2/σ^2 或 n 成正比。

对于信号的特征值不相等情况(例如非独立同分布信号)，不能得到式(3.2.94)的闭式表达式。此时，可以用近似值或临界值来分析检测性能，具体分析方法将在 3.3 节中介绍。

备注 3.2.4 独立与非独立信号模型间的关系

在式(3.2.85)中，当 $\boldsymbol{N}\sim\mathcal{N}(\boldsymbol{0},\ \sigma^2\boldsymbol{I})$，$\boldsymbol{S}\sim\mathcal{N}(\boldsymbol{\mu},\ \boldsymbol{\Sigma}_{\mathrm{s}})$ 且 $\boldsymbol{\Sigma}_{\mathrm{s}}=\mathrm{diag}\{\sigma_{s_1}^2,\ \cdots,\ \sigma_{s_n}^2\}$ 时，噪声样本具有独立同分布特性且服从 $\mathcal{N}(0,\ \sigma^2)$ 分布，信号也具有独立同分布特性且服从 $\mathcal{N}(\mu_k,\ \sigma_{s_k}^2)$ 分布。对应的对数似然比可以写为

$$\log L(\boldsymbol{y}) = \frac{1}{2}\sum_{k=1}^{n}y_k^2/\sigma^2-\frac{1}{2}\sum_{k=1}^{n}(y_k-\mu_k)^2/(\sigma_{s_k}^2+\sigma^2)+\frac{1}{2}\sum_{k=1}^{n}\log[\sigma^2/(\sigma_{s_k}^2+\sigma^2)] \tag{3.2.98}$$

下面进一步考虑 $\boldsymbol{\Sigma}_{\mathrm{s}}$ 不是对角阵时的情况。令 $p_j(y_1,\ \cdots,\ y_l)$ 表示 $Y_1,\ \cdots,\ Y_l$ 在 H_j 下的概率密度，则有

$$p_j(\boldsymbol{y}) = p_j(y_1)\prod_{k=2}^{n}p_j(y_k\,|\,y_1,\cdots,y_{k-1}) \tag{3.2.99}$$

其中 $p_j(y_k\,|\,y_1,\ \cdots,\ y_{k-1})$ 是 $Y_1=y_1,\ \cdots,\ Y_{k-1}=y_{k-1}$ 时 y_k 的条件概率密度。式(3.2.99)对 $\mathbf{R}^n$ 上的任意概率密度都成立，并且满足 $p_j(y_k\,|\,y_1,\ \cdots,\ y_{k-1})=p_j(y_1,\ \cdots,\ y_k)/p_j(y_1,\ \cdots,\ y_{k-1})$。在 H_0 假设下，由于 $\boldsymbol{N}$ 是独立同分布的，Y_k 不依赖于 $Y_1,\ \cdots,\ Y_{k-1}$，进而有 $p_0(\boldsymbol{y}) = \prod_{k=1}^{n}p_0(y_k)$ 成立。在 H_1 假设下，虽然 Y_k 与 $Y_1,\cdots,Y_{k-1}$ 不独立，但是因为 $\boldsymbol{Y}$ 是高斯随机向量，在给定 $Y_1=y_1,\cdots,Y_{k-1}=y_{k-1}$ 时，Y_k 服从条件高斯分布，对应的条件均值可以写为

$$\begin{aligned}
&E_1\{Y_k\,|\,Y_1=y_1,\cdots,Y_{k=1}=y_{k-1}\}\\
&=E_1\{S_k\,|\,Y_1=y_1,\cdots,Y_{k=1}=y_{k-1}\}\\
&\quad+E_1\{N_k\,|\,Y_1=y_1,\cdots,Y_{k=1}=y_{k-1}\}\\
&=E_1\{S_k\,|\,Y_1=y_1,\cdots,Y_{k=1}=y_{k-1}\}\triangleq\hat{S}_k
\end{aligned} \tag{3.2.100}$$

上式中利用了 N_k 不依赖于 $Y_1,\ \cdots,\ Y_{k-1}$ 且均值为 0 这一性质。类似地，条件方差为

$$\mathrm{var}_1(Y_k\,|\,Y_1=y_1,\cdots,Y_{k-1}=y_{k-1})=\mathrm{var}_1(S_k\,|\,Y_1=y_1,\cdots,Y_{k-1}=y_{k-1})$$

⊖ 注意，即使对于 n(n 为偶数)仍然可以对于式(3.2.86)进行分部积分并得到

$$P_j(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{Y}>\tau') = \mathrm{e}^{-\tau'/2\sigma_j^2}\sum_{k=0}^{n/2-1}\frac{(\tau'/2\sigma_j^2)^k}{k!}$$

$$+\mathrm{var}_1(N_k|Y_1=y_1,\cdots,Y_{k-1}=y_{k-1})=\hat{\sigma}^2_{S_k}+\sigma^2 \tag{3.2.101}$$

其中$\hat{\sigma}^2_{S_k}\triangleq\mathrm{var}_1(S_k|Y_1=y_1,\ \cdots,\ Y_{k-1}=y_{k-1})$。

多元高斯分布的一个特性是$\hat{\sigma}^2_{S_k}$不依赖于y_1，…，y_{k-1}的取值(该性质与其他相关性质将在随后几章中陆续介绍)。考虑$\hat{S}_1=E\{S_1\}$，$\hat{\sigma}^2_{S_1}=\mathrm{var}(S_1)$且$p_1(\boldsymbol{y})$是$\mathcal{N}(\hat{S}_k,\ \hat{\sigma}^2_{S_k}+\sigma^2)$密度的乘积，则对数似然比可表示为

$$\log L(\boldsymbol{y})=\frac{1}{2}\sum_{k=1}^{n}y_k^2/\sigma^2-\frac{1}{2}\sum_{k=1}^{n}(y_k-\hat{S}_k)^2/(\hat{\sigma}^2_{S_k}+\sigma^2)+\frac{1}{2}\sum_{k=1}^{n}\log[\sigma^2/(\hat{\sigma}^2_{S_k}+\sigma^2)] \tag{3.2.102}$$

对比式(3.2.102)与式(3.2.98)可以发现，检测一个非独立的统计信号与检测一个均值为$\boldsymbol{S}$、方差为$\mathrm{diag}\{\hat{\sigma}^2_{S_1},\ \cdots,\ \hat{\sigma}^2_{S_k}\}$的独立统计信号是相似的。不同之处在于式(3.2.102)中的$\hat{\boldsymbol{S}}_k$依赖于y_1，…，y_{k-1}，而式(3.2.98)中的μ_k则不是。

为了更好地理解上述结论，在假设H_1下将观测模型作如下修改

$$Y_k=N_k+S_k=N_k+\varepsilon_k+\hat{S}_k \tag{3.2.103}$$

其中，$\hat{S}_k=E_1\{S_k|Y_1,\ \cdots,\ Y_{k-1}\}$，$\varepsilon_k=S_k-\hat{S}_k$。在后续章节中将证明$\hat{S}_k$是$H_1$假设下利用观测数据$Y_1$，…，$Y_{k-1}$对$S_k$的最优预测，所以$\varepsilon_k$可解释为最优预测器的预测误差或$S_k$的不可预测部分。换言之，当前观测数据中的有用信号可分解为先前观测数据的可预测分量$\hat{S}_k$和不可预测分量ε_k[⊖]。可以证明，在H_1下，ε_k在统计意义上与Y_1，…，Y_{k-1}独立，且服从$\mathcal{N}(0,\ \hat{\sigma}^2_{S_k})$分布。通过对比可以发现，当$S_1$，…，$S_n$是服从$\mathcal{N}(\mu_k,\ \sigma^2_{S_k})$分布的独立序列时，$\hat{S}_k=\mu_k$且$\varepsilon_k=S_k-\mu_k$，其中$\varepsilon_k$服从$\mathcal{N}(0,\ \sigma^2_{S_k})$分布。

备注 3.2.5　独立同分布高斯噪声下最优统计检测的估计-相关解释

对式(3.2.102)的结构作进一步分析，可以得到$L(\boldsymbol{y})$如下

$$\log L(\boldsymbol{y})=\frac{1}{2\sigma^2}\left[\sum_{k=1}^{n}y_k^2-\sum_{k=1}^{n}(y_K-\hat{S}_k)^2/(1+\hat{\sigma}^2_{S_k}/\sigma^2)\right]-\frac{1}{2}\sum_{k=1}^{n}\log(1+\hat{\sigma}^2_{S_k}/\sigma^2) \tag{3.2.104}$$

假设噪声方差σ^2远大于最大预测误差的方差，即对所有k有$\hat{\sigma}^2_{S_k}\ll\sigma^2$成立。在这一条件下，$(1+\hat{\sigma}^2_{S_k}/\sigma^2)\approx1$成立，似然函数可改写为

$$\log L(\boldsymbol{y})\approx\frac{1}{\sigma^2}\left[\sum_{k=1}^{n}y_k\hat{S}_k-\frac{1}{2}\sum_{k=1}^{n}(\hat{S}_k)^2\right] \tag{3.2.105}$$

将上式与式(3.2.9)对比，可以发现式(3.2.105)是检测相干信号$\hat{S}_1$，…，$\hat{S}_n$的结构，即该情况下的统计信号检测器至少可近似为估计信号后的确知信号检测器。

在更一般的情况下，假设需要在服从$\mathcal{N}(\boldsymbol{0},\ \sigma^2\boldsymbol{I})$分布的噪声背景中检测不限于高斯分布的、概率密度函数为p_S的多变量统计信号，则式(3.2.85)的似然比为

$$L(\boldsymbol{y})=\int_{\mathbf{R}^n}\exp\left\{\frac{1}{\sigma^2}\left(\boldsymbol{s}^{\mathrm{T}}\boldsymbol{y}-\frac{1}{2}\|\boldsymbol{s}\|^2\right)\right\}p_{\boldsymbol{S}}(\boldsymbol{s})\mathrm{d}\boldsymbol{s} \tag{3.2.106}$$

⊖　有关$\hat{S}_k$和$N_k+\varepsilon_k$的类似表达方式也可以结合Y_k给出。例如，$\hat{S}_k$是Y_k中能由先前观察数据确定的分量，此外，$Y_k-\hat{S}_k=N_k+\varepsilon_k$也是$Y_k$的一部分，但不能由先前观察数据表示。也就是说，$Y_k-\hat{S}_k$包含了$Y_k$中的新信息。序列$Y_1-\hat{S}_1$，…，$Y_k-\hat{S}_k$通常被称为新息序列。这种表述方式在滤波中有着非常重要的作用，我们将在后续章节中详细分析。

在概率函数 p_S的正则性约束下，利用均值理论[Apostol(1974)]可以得到：

$$L(\boldsymbol{y}) = \exp\left\{\frac{1}{\sigma^2}\left(\hat{\boldsymbol{S}}^{\mathrm{T}}\boldsymbol{y} - \frac{1}{2}\|\hat{\boldsymbol{S}}\|^2\right)\right\} \tag{3.2.107}$$

其中$\hat{\boldsymbol{S}}$位于 $\mathbf{R}^n$ 空间($\hat{\boldsymbol{S}}$依赖于 $\boldsymbol{y}$)。由上式可以看出，一般来说稳态信号在独立同分布高斯噪声下的似然比可以表示为信号的估计器级联估计信号的最优检测器[⊖](也称为估计-相关结构)。不过，由于确定$\hat{\boldsymbol{S}}(\boldsymbol{y})$函数非常困难，因此该结构实际上较为复杂。此外，与式(3.2.105)不同，式(3.2.107)中的$\hat{\boldsymbol{S}}$通常依赖于所有观测数据 $\boldsymbol{y}$ 而无法实时计算。因此，估计-相关结构的意义更多是用来指导设计独立同分布高斯噪声中的统计信号次优检测器。例如，可以首先构建信号的估计器，然后将该估计作为一个已知信号。在连续时间模拟信号检测中，估计-相关结构会以更为直接的特征估计器的形式出现，具体内容将在第 6 章作进一步说明。

备注 3.2.6　统计信号的局部最优检测

式(3.2.86)所描述的二次型检测器在某些情况下不具有一致最大势特性。例如，考虑下面的检测模型

$$\begin{aligned} H_0: Y_k &= N_k, & k &= 1,2,\cdots,n \\ H_1: Y_k &= N_k + \theta^{1/2} S_k, & k &= 1,2,\cdots,n, \quad \theta > 0 \end{aligned} \tag{3.2.108}$$

其中，$\boldsymbol{N}$ 和 $\boldsymbol{S}$ 与式(3.2.85)相同且具有单位噪声方差。在这种情况下，$\theta^{1/2}\boldsymbol{S}$ 的方差变为 $\theta\boldsymbol{\Sigma}_{\mathrm{s}}$。给定 θ 后，尼曼-皮尔逊的二次型检测统计可以表示为

$$\theta\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{s}}(\boldsymbol{I} + \theta\boldsymbol{\Sigma}_{\mathrm{s}})^{-1}\boldsymbol{y} \tag{3.2.109}$$

尽管式(3.2.109)前面的系数 θ 可以并入检测阈值，但是无法消除$(\boldsymbol{I}+\theta\boldsymbol{\Sigma}_{\mathrm{s}})^{-1}$中 θ 与观测数据 $\boldsymbol{y}$ 的耦合。所以式(3.2.108)所描述的统计量不是 UMP(一致最大势)统计量。不过，可以将式(3.2.109)中的 θ 设为 0，得到 LMP(局部最大势)统计量为

$$2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{s}}\boldsymbol{y} \tag{3.2.110}$$

当 $\boldsymbol{\Sigma}_{\mathrm{s}}$ 的第 $k-l$ 个元素 $\rho_{k,l}$仅依赖于 $k-l$ 的大小时，式(3.2.110)给出的统计量有一种非常有趣的解释。此时，为方便起见，忽略第二个下标后，有 $\rho_{k,l}=\rho_{k-l,0}\triangleq\rho_{k-l}$成立。把满足上述特征的信号称为广义平稳信号，具体概念将在第 5 章进行详细说明。考虑比例变换后的 LMP 统计量为

$$T(\boldsymbol{y}) = \frac{1}{n}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathrm{s}}\boldsymbol{y} \tag{3.2.111}$$

由于比例变换仅仅对阈值作了改动，因此，上式与式(3.2.110)等价。通过简单的数学运算(见习题 18)后，$T(\boldsymbol{y})$可以写为

$$T(\boldsymbol{y}) = \frac{1}{n}\sum_{k=1}^{n}\sum_{l=1}^{n} y_k y_l \rho_{k-l} = \rho_o\,\hat{\rho}_o + 2\sum_{k=1}^{n-1}\rho_k\,\hat{\rho}_k \tag{3.2.112}$$

其中$\hat{\rho}_k$ 定义为

⊖ 该描述对任意噪声密度 p_N 都是有效的，其中，p_N 对于每一个 $y\in\mathbf{R}^n$ 足够光滑(存在微分)，$\hat{\boldsymbol{S}}\in\mathbf{R}^n$

$$L(\boldsymbol{y}) = \int_{\mathbf{R}^n}\frac{p_N(\boldsymbol{y}-\boldsymbol{s})}{p_N(\boldsymbol{y})}p_S(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} = \frac{p_N(\boldsymbol{y}-\hat{\boldsymbol{S}})}{p_N(\boldsymbol{y})}$$

$$\hat{\rho}_k \triangleq \frac{1}{n-k}\sum_{l=1}^{n-k} y_l y_{l+k}, \quad k=0,\cdots,n-1 \tag{3.2.113}$$

结合式(3.2.112)的形式，考虑到 $n \gg k$ 时，$\hat{\rho}_k$ 是协方差 $E\{Y_l Y_{l+k}\}$（$l=1$，…，$n-k$）的估计，则式(3.2.111)中 LMP 统计量 $T(\mathbf{y})$可解释为估计观测数据的协方差矩阵后与信号协方差序列相关。在 H_0 假设下，有

$$E\{Y_l Y_{l+k}\} = \begin{cases} 1, & k=0 \\ 0, & k \neq 0 \end{cases} \tag{3.2.114}$$

相应地，在 H_1 下，有

$$E\{Y_l Y_{l+k}\} = \begin{cases} 1+\theta\rho_0, & k=0 \\ \theta\rho_k, & k \neq 0 \end{cases} \tag{3.2.115}$$

因此，假设估计值$\hat{\rho}_k$ 足够精确，有下面的表达式成立

$$T(\mathbf{y}) \approx \begin{cases} \rho_0, & H_0 \\ \rho_0 + \theta(\rho_0^2 + 2\sum_{k=1}^{n-1}\rho_k^2), & H_1 \end{cases} \tag{3.2.116}$$

从上式可以看出，统计量 $T(\mathbf{y})$是直观合理的，特别是当信号高度相关时（即 $\sum_{k=1}^{n}\rho_k^2$ 的值非常大）。

在频域中，统计量 $T(\mathbf{y})$同样也有另一种解释。特别地，假设 S_1，…，S_n 是无限长序列$\{S_k\}_{k=-\infty}^{\infty}$的一个子集，其中假设序列$\{S_k\}_{k=-\infty}^{\infty}$对于每一个整数 l 和 k 都有 $E\{S_l S_{l+k}\}=\rho_k$ 成立，则$\{\rho_k\}_{k=-\infty}^{\infty}$的 DFT 可以定义为

$$\phi(\omega) \triangleq \sum_{k=-\infty}^{\infty}\rho_k e^{i\omega k} \tag{3.2.117}$$

上式是$\{S_k\}_{k=-\infty}^{\infty}$ 的功率谱。代入 $\rho_k = \frac{1}{2\pi}\int_{-\pi}^{\pi}\phi(\omega)e^{-i\omega k}d\omega$ 后，$T(\mathbf{y})$ 可以写为

$$T(\mathbf{y}) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\phi(\omega)\,\hat{\phi}(\omega)d\omega \tag{3.2.118}$$

其中

$$\hat{\phi}(\omega) \triangleq \frac{1}{n}\left|\sum_{k=1}^{n} y_k e^{i\omega k}\right|^2, \quad -\pi \leqslant \omega \leqslant \pi \tag{3.2.119}$$

函数$\hat{\phi}$也称为数据的周期图，它是对观测数据谱的一种估计。因此，在式(3.2.118)中，$T(\mathbf{y})$首先估计观测数据的谱，然后在频域将估计值与信号谱进行相关。在 H_0 下，观测数据的谱在 $\omega\in[-\pi, \pi]$内恒为 1；在 H_1 下，观测数据的谱在 $\omega\in[-\pi, \pi]$内为 $1+\phi(\omega)$。因此，式(3.2.118)与式(3.2.112)有类似的解释。

3.3 信号检测算法的性能评估

3.2 节对时间离散信号的最优检测算法进行了设计和分析。在某种意义上，设计这些算法比分析它们的性能来得更直接，因为性能分析通常较为棘手。之所以能够计算 3.2 节中大多数示例的性能，仅仅是因为所采用的模型易于处理，因此将其作为说明这些检测理论的最佳模型。在一些情况下，仅对上述示例中的假设做一些微小的改变，也会大大增加性能分析的难度。例如，在独立同分布的高斯噪声环境下检测高斯信号时，如果信号也是

独立同分布的，则计算误差概率是非常容易的；但如果只要有一个信号样本的方差发生变化，计算误差概率就变得十分困难。

给定判决准则$\widetilde{\delta}$，则二元假设检验问题的基本性能由两个条件误差概率P_{F}和P_{M}决定，分别定义为

$$P_{\mathrm{F}}(\widetilde{\delta}) = P_0(\widetilde{\delta}\text{选择 } H_1)$$
$$P_{\mathrm{M}}(\widetilde{\delta}) = P_1(\widetilde{\delta}\text{选择 } H_0)$$

似然比检测和其他大多数判决准则的形式都可以写成

$$\widetilde{\delta}_T(y) = \begin{cases} 1, & T(y) > \tau \\ \gamma, & T(y) = \tau \\ 0, & T(y) < \tau \end{cases} \tag{3.3.1}$$

式中，T是一个从$(\Gamma, \mathcal{G})$到$(\mathbf{R}, \mathcal{B})$的映射（比如对数似然比）。因此，分析系统性能需要分别计算在两种假设下区间$\{T(Y)>\tau\}$（或$\{T(Y)<\tau\}$）和$\{T(Y)=\tau\}$的概率。虽然计算概率在概念上很简单，但在实际中往往难以得到解析式。比如，假设$Y=\{Y_1, \cdots, Y_n\}$在H_0下的联合概率分布函数为p_0，则

$$P_{\mathrm{F}}(\widetilde{\delta}_T) = \int\cdots\int_{T(Y)>\tau} p_0(y_1,\cdots,y_n)\mathrm{d}y_1\cdots\mathrm{d}y_n + \gamma\int\cdots\int_{T(Y)=\tau} p_0(y_1,\cdots,y_n)\mathrm{d}y_1\cdots\mathrm{d}y_n$$

在n值比较大时，如果不对上式进行进一步简化，计算就变得非常困难。在本节中，讨论几种精确计算检测性能、估计检测性能下界以及近似计算检测性能的常用方法。

3.3.1 直接计算检测性能

结合3.3.1节中的系统模型，有

$$\begin{aligned} P_{\mathrm{F}}(\widetilde{\delta}_T) &= P(T(Y) > \tau | H_0) + \gamma P(T(Y) = \tau | H_0) \\ &= [1 - F_{T,0}(\tau)] + \gamma[F_{T,0}(\tau) - \lim_{\sigma\to\tau^-} F_{T,0}(\sigma)] \end{aligned}$$

和

$$\begin{aligned} P_{\mathrm{M}}(\widetilde{\delta}_T) &= P(T(Y) < \tau | H_1) + (1-\gamma)P(T(Y) = \tau | H_1) \\ &= P(T(Y) \leqslant \tau | H_1) - \gamma P(T(Y) = \tau | H_1) \\ &= F_{T,1}(\tau) - \gamma[F_{T,1}(\tau) - \lim_{\sigma\to\tau^-} F_{T,1}(\sigma)] \end{aligned}$$

其中，$F_{T,j}$是假设H_j下$T(Y)$的累计分布函数。因此，对于3.3.1节所描述的检测器形式（如似然比检测器），如果能够较容易地确定在阈值τ邻域内$F_{T,j}$的累计分布函数，则分析检测性能会相对简单。

当$\mathbf{Y}=(Y_1,\cdots,Y_n)^{\mathrm{T}}$是独立（实）随机变量，且$T(\mathbf{y})=\sum_{k=1}^{n} g_k(y_k)$时，$F_{T,j}$具有简单表达形成。其中，$\{g_k\}_{k=1}^n$是非线性序列（例如对数似然比）。此时，利用特征函数来计算$F_{T,j}$就较为容易。随机变量X的特征函数定义为

$$\phi_X(u) = E\{\mathrm{e}^{\mathrm{i}uX}\}, \quad u \in \mathbf{R}$$

其中，i表示虚数单位$\sqrt{-1}$。注意，累计分布函数F和它的特征函数ϕ一一对应[⊖]。用

⊖ 有关特征方程的问题在Lukacs(1960)中进行了详细的讨论。

$\phi_{T,j}$和$\phi_{g_k,j}$分别表示在H_j假设下$T(\mathbf{Y})$和$g_k(Y_k)$的特征函数，根据Y_k的独立性可得

$$\phi_{T,j}(u)=E\left\{\exp\left\{\mathrm{i}u\sum_{k=1}^{n}g_k(Y_k)\right\}\mid H_j\right\}$$

$$=\prod_{k=1}^{n}E\{\exp\{\mathrm{i}ug_k(Y_k)\}\mid H_j\}=\prod_{k=1}^{n}\phi_{g_k,j}(u)$$

对特征函数$\phi_{T,j}$进行逆变换可得到累积分布$F_{T,j}$[具体可参见 Billingsley(1979)]

$$F_{T,j}(b)-F_{T,j}(a)=\lim_{U\to\infty}\frac{1}{2\pi}\int_{-U}^{U}\frac{\mathrm{e}^{-\mathrm{i}ua}-\mathrm{e}^{-\mathrm{i}ub}}{\mathrm{i}u}\phi_{T,j}(u)\mathrm{d}u \tag{3.3.2}$$

上式对所有的a和b都是成立的，其中a和b是$F_{T,j}$的连续点。根据$F_{T,j}(b)-F_{T,j}(a)$在所有连续点上的值足以唯一确定$F_{T,j}$，因为$F_{T,j}(-\infty)=0$，而且$F_{T,j}$必须是右连续的。如果$T(\mathbf{Y})$在H_j假设下是连续的随机变量，则式(3.3.2)的反变换十分简单，因为$F_{T,j}$的概率密度函数$p_{T,j}$可以表示为

$$p_{T,j}(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\phi_{T,j}(u)\mathrm{e}^{-\mathrm{i}ut}\mathrm{d}t$$

显然，$p_{T,j}$和$\phi_{T,j}$是傅里叶变换对[⊖]。此时，最终能得到

$$P_{\mathrm{F}}(\tilde{\delta}_T)=\frac{1}{2\pi}\int_{\tau}^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}ut}\left[\prod_{k=1}^{n}\phi_{gk,o}(u)\right]\mathrm{d}u\mathrm{d}t \tag{3.3.3}$$

同理，可以得到$P_{\mathrm{M}}(\tilde{\delta}_T)$的计算表达式。

例 3.3.1 柯西噪声下的相关检测

作为上述方法的例子，考虑在加性柯西噪声下利用相关检测器(定义为$g_k(y_k)=s_ky_k$，$k=1$，…，n)来检测相干信号的性能。特别地，考虑假设检验对如下

$$H_0:Y_k=N_k,\qquad k=1,\cdots,n$$

$$H_1:Y_k=N_k+s_k,\quad k=1,\cdots,n$$

其中，N_1，…，N_n是独立同分布的随机变量序列，且服从柯西分布，概率密度为

$$p_{N_k(x)}=\frac{1}{\pi(1+x^2)},\quad x\in\mathbf{R}$$

s_1，…，s_n是已知的信号序列。在这种情况下，有

$$\phi_{g_k,0}(u)=E\{\mathrm{e}^{\mathrm{i}us_kN_k}\}=\phi_{N_k}(us_k)=F\{p_{N_k}\}\mid_{us_k}=\mathrm{e}^{|-s_ku|},\quad u\in\mathbf{R}$$

因此

$$\phi_{T,0}(u)=\prod_{k=1}^{n}\phi_{g_k,0}(u)=\mathrm{e}^{-n|\bar{s}||u|}$$

其中，$|\bar{s}|\triangleq(1/n)\sum_{k=1}^{n}|s_k|$。由于$\phi_{T,0}(u)$绝对可积，有

$$P_{\mathrm{F}}(\tilde{\delta}_T)=\frac{1}{2\pi}\int_{\tau}^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}ut}\mathrm{e}^{-n|\bar{s}||u|}\mathrm{d}u\mathrm{d}t=\frac{1}{n|\bar{s}||u|}\int_{\tau}^{\infty}\frac{1}{1+(t/n|\bar{s}|)^2}\mathrm{d}t$$

$$=\frac{1}{2}-\frac{1}{\pi}\arctan\left(\frac{\tau}{n|\bar{s}|}\right)$$

⊖ 在H_j假设下，$T(\mathbf{Y})$的充分(非必要)条件是$\int_{-\infty}^{\infty}|\phi_{T,j}(u)|\mathrm{d}u<\infty$[参见 Breiman(1968)]。

类似地，有 $\phi_{T,1}(u)=\phi_{T,0}(u)\mathrm{e}^{\mathrm{i}un\overline{s^2}}$，其中 $\overline{s^2}\triangleq(1/n)\sum_{k=1}^{n}s_k^2$，进而可直接得到

$$P_{\mathrm{M}}(\widetilde{\delta}_T)=\frac{1}{2}+\frac{1}{\pi}\arctan\left(\frac{\tau}{n\,|s|}-\sqrt{\overline{s^2}}\right)$$

[值得注意的是，通过选择 $\tau=n\,|\bar{s}|\tan(1/2-\alpha)$ 可以确保虚警概率为 α。相应的 ROC 为

$$P_{\mathrm{D}}(\widetilde{\delta}_T)=1-P_{\mathrm{M}}(\widetilde{\delta}_T)=\frac{1}{2}-\frac{1}{\pi}\arctan\left[\tan\left(\frac{1}{2}-\alpha\right)-\sqrt{\overline{s^2}}\right]\tag{3.3.4}$$

式(3.3.4)中隐含了一个结论：仅通过提高信号平均功率就能够改善检测性能，而增加信号样本不能改善检测性能。诚然，通过增加样本确实能提高最优检测器的性能，但在柯西噪声背景下检测相干信号时，相关检测器与最优检测器存在明显差异。从本质上讲，柯西分布的长拖尾抵消了相关检测器在减少噪声方面的作用。]　◀

在例 3.3.1 中，利用特征函数方法同样能得到 P_{F} 和 P_{M} 的闭式解。然而，特征函数法不是总能得到易于处理的误差概率闭式解，通常需要以式(3.3.3)和 P_{M} 的相应表达式为基础来近似计算误差概率。就一般的误差概率值而言，采用快速傅里叶变换算法(FFT)等方法对相应的 P_{M} 和式(3.3.3)进行数值积分后，可得到高精度的近似结果。对于误差概率均很小的情况($\leqslant 10^{-5}$)，需要利用其他更有效的数值计算方法。当某个参数(阈值、信噪比、样本个数等)非常大时，可利用与该参数有关的误差概率渐近表达式来近似计算误差概率。在通信系统分析中广泛应用的*鞍点近似*就是利用渐近表达式计算误差概率的典型示例[详见 DeBruijn(1961)]。具体应用示例可参见 Lugannani 和 Rice(1980)的文章，他们给出了 $g_k(Y_k)$ 同分布下的近似方法；Mazo 和 Salz(1965)在文章中利用*鞍点近似*分析了二次型检测器(见 3.2 节)在非独立同分布高斯信号和噪声下的性能。

3.3.2　切诺夫界

精确计算 3.3.1 节中检测器的误差概率 P_{F} 和 P_{M} 通常是不现实(或者是不可能)的。不过，就某些应用而言，若能获得误差概率的准确上界就已经足够。在本书中使用切诺夫界，它是似然比检测器性能的边界，可由马尔可夫不等式推导出来。

马尔可夫不等式　*假设 X 是随机变量，如果 $P(X\geqslant 0)=1$，则对于所有的 $a>0$，有 $P(X\geqslant a)\leqslant E\{X\}/a$ 成立。*

证明　$P(X\geqslant a)\leqslant E\{I_{[a,\infty)}(X)\}$，其中 $I_{[a,\infty)}$ 是集合$[a,\infty)$上的指示函数，定义为

$$I_{[a,\infty)}(X)=\begin{cases}1, & x\geqslant a\\ 0, & x<a\end{cases}$$

对于 $X\geqslant 0$，有 $I_{[a,\infty)}(X)\leqslant X/a$，代入后可完成马尔可夫不等式的证明。　□

考虑随机量 $\gamma\leqslant 1$，并将马尔可夫不等式应用到 3.3.1 节中，可得到误差概率的计算式为

$$\begin{aligned}P_{\mathrm{F}}(\widetilde{\delta}_T)&\leqslant P_0(T(Y)\geqslant\tau)=P_0(\mathrm{e}^{sT(Y)}\geqslant\mathrm{e}^{s\tau})\leqslant\mathrm{e}^{-s\tau}E\{\mathrm{e}^{sT(Y)}\mid H_0\}\\&=\exp\{-s\tau+\mu_{T,0}(s)\}\end{aligned}\tag{3.3.5}$$

其中，$s>0$ 和 $\mu_{T,0}$ 是 $T(Y)$ 在 H_0 假设下的累积量母函数，定义为

$$\mu_{T,0}(s)=\log E\{\mathrm{e}^{sT(Y)}\mid H_0\}$$

类似地，在 $\gamma\geqslant 0$ 时，有

$$P_{\mathrm{M}}(\tilde{\delta}_T) \leqslant P_1(T(Y) \leqslant \tau) = P_1(\mathrm{e}^{tT(Y)} \geqslant \mathrm{e}^{t\tau}) \leqslant \exp\{-t\tau + \mu_{T,1}(t)\} \quad (3.3.6)$$

其中，$t<0$ 和 $\mu_{T,1}$ 是 $T(Y)$ 在 H_1 假设下的累积量母函数。

假设 $T(Y)$ 的累积量母函数已知，在 $s>0$ 和 $t<0$ 时对式(3.3.5)和式(3.3.6)最小化可得到紧边界。该边界对似然比检测器来说非常有用。为了说明这一问题，假设在 $j=0$ 和 1 时 P_j 的概率密度为 p_j，并且选择 $T(y)=\log L(y)$，其中 $L=p_1/p_0$。此时，有

$$\mu_{T,0}(s) = \log\left(\int_\Gamma \mathrm{e}^{s\log L} p_0 \mathrm{d}\mu\right) = \log\left(\int_\Gamma L^s p_0 \mathrm{d}\mu\right)$$

和

$$\mu_{T,1}(t) = \log\left(\int_\Gamma L^t p_1 \mathrm{d}\mu\right) = \log\left(\int_\Gamma L^{t+1} p_0 \mathrm{d}\mu\right) = \mu_{T,0}(t+1)$$

成立。代入式(3.3.6)并进行变量替换后，有

$$P_{\mathrm{M}}(\delta_T) \leqslant \exp\{(1-s)\tau + \mu_{T,0}(s)\}, \quad s<1 \quad (3.3.7)$$

需要注意的是，如果

$$\arg\{\min_{s<1}[\mu_{T,0}(s) - s\tau]\} > 0$$

和

$$\arg\{\min_{s>0}[\mu_{T,0}(s) - s\tau]\} < 1$$

成立，则式(3.3.5)和式(3.3.7)可以在 s 值相同时达到最小。可以证明，在 $\mu_{T,0}(s)<\infty$ 区间上 $(\mu_{T,0}(s)-s\tau)$ 是 s 的凸函数[参见 Billingsley(1979)]。因此，为使 $[\mu_{T,0}(s)-s\tau]$ 达到最小，以下条件量充分的

$$\mu'_{T,0}(s) = \tau \quad (3.3.8)$$

其中 $\mu'_{T,0}(s)=\mathrm{d}\mu_{T,0}(s)/\mathrm{d}s$。同样可以证明，对于 $j=0$ 和 1，如果 $E\{|\log L(y)||H_j\}<\infty$ 成立，则

$$\mu'_{T,0}(j) = E\{\log L(y) | H_j\} \quad (3.3.9)$$

又因为 $(\mu_{T,0}(s)-s\tau)$ 是凸函数，即 $\mu'_{T,0}(s)$ 是 s 的增函数，则式(3.3.9)在 $s\in(0, 1)$ 内有解的条件是当(且仅当)下式成立

$$\mu_0 \triangleq E\{\log L(Y) | H_0\} < \tau < E\{\log L(Y) | H_1\} \triangleq \mu_1 \quad (3.3.10)$$

因此，假设式(3.3.10)成立，式(3.3.5)和式(3.3.7)可改写成

$$P_{\mathrm{F}}(\tilde{\delta}_T) \leqslant \exp\{\mu_{T,0}(s_0) - s_0\mu'_{T,0}(s_0)\} \quad (3.3.11)$$

$$P_{\mathrm{M}}(\tilde{\delta}_T) \leqslant \exp\{\mu_{T,0}(s_0) + (1-s_0)\mu'_{T,0}(s_0)\} \quad (3.3.12)$$

其中

$$\mu_0 < \mu'_{T,0}(s_0) = \tau < \mu_1 \quad (3.3.13)$$

式(3.3.11)和式(3.3.12)即为**切诺夫界**。因为 $(\mu_{T,0}(s)-s\tau)$ 是凸的，所以在 $\tau\leqslant\mu_0$ 时，$\min_{s\geqslant 0}[\mu_{T,0}(s)-s\tau]=0$ 表明式(3.3.5)的上界无意义(即 $\geqslant 1$)，在 $\tau\geqslant\mu_1$ 时式(3.3.7)的边界也无意义。当然，其他形式的上界可能是有意义的。需要注意的是，Jensen 不等式[⊖]表明只有 $P_0=P_1$ 时，不等式 $\mu_0\leqslant 0$ 和 $\mu_1\geqslant 0$ 取等号，因此阈值 $\tau=0$ 总能满足式(3.3.10)。

⊖ Jensen 不等式：对于任意变量 X 和凸函数 C，有 $E\{C(X)\}\geqslant C(E\{X\})$。当 C 是严格凸函数，且 $P(X=E\{X\})=1$ 时，等号成立。

如果先验信息 π_0 和 π_1 已知，那么根据式(3.3.11)和式(3.3.12)能够得到平均误差概率的上界。特别地

$$P_e = \pi_0 P_F + \pi_1 P_M \leqslant [\pi_0 + \pi_1 e^{\mu'_{T,0}(s_0)}]\exp\{\mu_{T,0}(s_0) - s_0\mu'_{T,0}(s_0)\} \tag{3.3.14}$$

不过，考虑到(见习题20)

$$P_e \leqslant \pi_0 e^{-s\tau}\int_{\Gamma_1} L^s p_0 \mathrm{d}\mu + \pi_1 e^{(1-s)\tau}\int_{\Gamma_0} L^s p_0 \mathrm{d}\mu \tag{3.3.15}$$

其中，$0\leqslant s\leqslant 1$，$\Gamma_1=\{L(Y)\geqslant\tau\}$，$\Gamma_0=\Gamma_1^c$，能够得到 P_e 更紧致的边界。式(3.3.15)表明

$$P_e \leqslant \max\{\pi_0, \pi_1 e^{\tau}\}\exp\{\mu_{T,0}(s) - s\tau\}, \quad 0\leqslant s\leqslant 1 \tag{3.3.16}$$

代入式(3.3.13)的条件后，式(3.3.16)的边界可表示为

$$P_e \leqslant \max[\pi_0, \pi_1 e^{\mu'_{T,0}(s_0)}] \times \exp\{\mu_{T,0}(s_0) - s_0\mu'_{T,0}(s_0)\} \tag{3.3.17}$$

除非 π_0 等于0或1，否则式(3.3.17)是比式(3.3.14)更紧的边界。对于阈值为 $\tau=\log(\pi_0/\pi_1)$ 的最小化误差概率的对数似然比检测器而言，式(3.3.16)可化简为

$$P_e \leqslant \pi_0^{1-s}\pi_1^s e^{\mu_{T,0}(s)}, \quad 0\leqslant s\leqslant 1 \tag{3.3.18}$$

特别地，在 $\Gamma=\mathbf{R}^n$ 时，$Y=(Y_1, \cdots, Y_n)$ 是一个独立同分布的观测序列，且在 H_j 假设下的边缘密度为 f_j 时，有下式成立

$$\mu_{T,0}(s) = n\log\left(\int_{\mathbf{R}} f_1^s f_0^{1-s}\right)$$

依据Jensen不等式，对于 $0<s<1$ 有下式成立

$$\log\left(\int_{\mathbf{R}} f_1^s f_0^{1-s}\right) < 0$$

此时，依据式(3.3.18)，误差概率随着样本数(n)的增加呈指数下降。

例3.3.2 平方律检波的切诺夫界

下面以在 $\mathcal{N}(\mathbf{0}, \sigma^2\boldsymbol{I})$ 分布的噪声中检测 $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_s)$ 分布的信号为例，说明切尔诺夫边界在分析检测器性能方面的作用。利用式(3.2.87)的方法将序列 $Y_1, \cdots, Y_n$ 变换成序列 $\overline{Y}_1, \cdots, \overline{Y}_n$ 后，有

$$L(\mathbf{y}) = \prod_{k=1}^{n}\frac{\sigma_{0k}}{\sigma_{1k}}e^{\bar{y}_k^2}$$

其中 σ_{jk}^2 参见式(3.2.90)。代入 $T=\log L$ 后，有

$$\mu_{T,0}(s) = \log\left(E\left\{\prod_{k=1}^{n}\left(\frac{\sigma_{0k}}{\sigma_{1k}}\right)^s e^{s\overline{Y}_k^2} \mid H_0\right\}\right) = \sum_{k=1}^{n} s\log\left(\frac{\sigma_{0k}}{\sigma_{1k}}\right) + \sum_{k=1}^{n}\log(E\{e^{s\overline{Y}_k^2} \mid H_0\}) \tag{3.3.19}$$

式(3.3.19)中的期望可以表示为

$$E\{e^{s\overline{Y}_k^2} \mid H_0\} = \begin{cases}[1 - s\sigma_{0k}^2]^{-1/2}, & s < 1/\sigma_{0k}^2 \\ \infty, & s \geqslant 1/\sigma_{0k}^2\end{cases}$$

既然对所有的 k，有 $\sigma_{0k}^2<1$，且当 $s\leqslant 1$ 时式(3.3.19)有界，则最小误差概率的边界可表示为

$$P_e \leqslant \pi_0^{1-s}\pi_1^s\prod_{k=1}^{n}\frac{\sigma^s(\sigma^2+\lambda_k)^{(1-s)/2}}{[\sigma^2+(1-s)\lambda_k]^{1/2}}, \quad 0\leqslant s\leqslant 1 \tag{3.3.20}$$

通过求解满足下式的 s_0 值，可以得到误差概率边界的最小值，即

$$2\log(\pi_0/\pi_1)+\sum_{k=1}^{n}\log(1+\lambda_k/\sigma^2)=\sum_{k=1}^{n}\frac{\lambda_k}{\sigma^2+(1-s_0)\lambda_k} \tag{3.3.21}$$

对于独立同分布情况($\lambda_1=\cdots=\lambda_n$)，可以直接求解式(3.3.21)。至于其他情况，则可通过数值计算方法求解。 ◀

与切诺夫界类似，在实际应用中还有其他几种非常有用的界。特别地，在式(3.3.18)中令 $s=1/2$ 后，可得到

$$P_e\leqslant\sqrt{\pi_0\pi_1}\exp\left\{\mu_{T,0}\left(\frac{1}{2}\right)\right\} \tag{3.3.22}$$

这种边界有时也称为 Bhattacharyya 边界，Bhattacharyya 系数为

$$\rho\triangleq\exp\left\{\mu_{T,0}\left(\frac{1}{2}\right)\right\}=\int_\Gamma[p_0p_1]^{1/2}\mathrm{d}\mu$$

这一系数又称 Hellinger 积分(或近似度积分)。根据系数 ρ，也可以得到 P_e 的下界。例如，有[参见 Kobayashi 和 Thomas(1967)]

$$\pi_0\pi_1\rho^2\leqslant P_e\leqslant(\pi_0\pi_1)^{1/2}\rho \tag{3.3.23}$$

结合 J 散度的定义

$$J=\int_\Gamma(L-1)\log(L)p_0\mathrm{d}\mu$$

可得到误差概率的下界为[Kobayashi 和 Thomas(1967)]

$$P_e>\pi_0\pi_1\mathrm{e}^{-J/2}$$

[J 的值与 p_0 和 p_1 之间的相对熵密切相关，参见 Kullback(1959)]相比 P_e 而言，计算 ρ 和 J 有时会更加容易，因此它们常作为选择信号的准则或其他相关准则的基础(例如选择 p_0 或 p_1)。

之所以采用切诺夫和上述其他边界，是因为计算 $E\{L^s\}$和 $E\{(L-1)\log L\}$通常比计算误差概率要容易些，且在边界上，函数 L 一致大于误差概率对应的指示函数。换句话说，如果想求解 $E\{h(L)\}$的边界，其中 h 是决策区域的指示函数，则可以寻找 $g\geqslant h$ 的函数使得 $E\{g(L)\}$更容易计算[㊀]。事实上，矩空间的一类误差概率边界均通过相似准则来计算，在这个边界中，试图寻找一个函数 g 使得 $E\{g(L)\}$易于计算，且$(h(L),g(L))$与 L 的变化曲线可近似为一条直线。这类边界由 Yao 和 Tobin(1976)提出，在评估包括符号间干扰和多址接入等噪声类型对数字通信系统的性能影响方面非常有用。

3.3.3 渐近相对有效性

在 3.3.1 节和 3.3.2 节中，研究了信号检测中直接计算误差概率和确定其上界的方法。不过，一般而言，精确获得误差概率较为困难。为此，需要有一些比误差概率更容易计算的性能指标。例如，渐近相对有效性(AER)就是在大样本、弱信号条件下衡量离散时间检测系统的有效准则[㊁]。对 AER 的推导如下。

假设有观测数据 Y_1，Y_2，…，服从两种统计假设 H_0 和 H_1 中的一种，再进一步假设

㊀ Boekee 和 Ruitenbeck(1981)对这类上下界进行了研究。

㊁ 在实际应用中，信号被噪声淹没，此时出现大样本数和微弱信号情况(例如被动声呐和雷达)，这就需要对信号进行长时间的相干积累。事实上，ARE 有很多种定义，但是最常用的是采用 E. J. G. Pitman 给出定义方法[见 Noether(1955)]。

$\widetilde{\delta}_1$ 和 $\widetilde{\delta}_2$ 是 H_0 相对于 H_1 的两个不同判决，它们有相同的误差概率，但使用的样本个数分别为 n_1 和 n_2。如果 $n_1<n_2$，说明 $\widetilde{\delta}_1$ 比 $\widetilde{\delta}_2$ 更有效，因为在相同的性能下 $\widetilde{\delta}_1$ 需要的信息比 $\widetilde{\delta}_2$ 少。类似地，如果 $n_1>n_2$，说明 $\widetilde{\delta}_2$ 比 $\widetilde{\delta}_1$ 有效。因此，n_2/n_1 的比值能衡量 $\widetilde{\delta}_1$ 相对于 $\widetilde{\delta}_2$ 的有效性。这种计算相对有效性的方法在大样本条件下(n_1 和 n_2 较大)非常有用，因为即使两个判决准则在大样本数下都具有优良的误差概率特性，如果样本数之比 $n_2/n_1=2$，同时 n_1 很大且 $\widetilde{\delta}_1$ 和 $\widetilde{\delta}_2$ 有相同的计算复杂度，那么 $\widetilde{\delta}_1$ 比 $\widetilde{\delta}_2$ 更可取。皮特曼定义的 ARE 是检测器效率的渐近(n_1，$n_1\to\infty$)计算，即在实现同等性能时一个检测器相对于另一个检测器所需的相对样本数。

为了更加准确地定义 ARE 这个概念，考虑 H_0 相对于 H_1 的两组判决序列 $\{\widetilde{\delta}_1^{(n)}\}_{n=1}^{\infty}$ 和 $\{\widetilde{\delta}_2^{(n)}\}_{n=1}^{\infty}$，其中 $\widetilde{\delta}_j^{(n)}$ 对应样本数为 n 的判决。假设每组序列的每个判决的虚警概率 $\alpha\in(0,1)$ 固定。对于样本数 n，$\{\widetilde{\delta}_1^{(n)}\}_{n=1}^{\infty}$ 相对于 $\{\widetilde{\delta}_2^{(n)}\}_{n=1}^{\infty}$ 的有效性定义为 n_2/n，其中 n_2 是满足 $P_D(\widetilde{\delta}_2^{(n_2)})\geqslant P_D(\widetilde{\delta}_1^{(n)})\triangleq\beta_n$ 的最小样本数。进而定义 ARE 为 n 趋向于无穷时相对有效性的极限。对于大多数合理的判决序列而言，几乎总有 $\lim\limits_{n\to\infty}\beta_n=1$ [⊖] 成立，即 n 非常大时，β_n 不再是衡量某个判决优于另一个判决的合适标准。为了解决这一难题，考虑另一个假设序列 $\{H_1^{(n)}\}_{n=1}^{\infty}$ 以某种方式收敛至 H_0，即 $\lim\limits_{n\to\infty}\beta_n\triangleq\beta\in(\alpha,1)$。假设 $\widetilde{\delta}_1^{(n)}$ 和 $\widetilde{\delta}_2^{(n)}$ 都是 H_0 对 $H_1^{(n)}$ 的判决，则可计算相对有效性，且 $\{\widetilde{\delta}_1^{(n)}\}_{n=1}^{\infty}$ 的渐近有效性定义为

$$\mathrm{ARE}_{1,2}=\lim_{n\to\infty}(n_2/n)$$

注意，$H_1^{(n)}$ 趋向于 H_0(当 n 较大时)，这与 2.5 节讨论的局部检测相吻合(例如弱信号检测)。

一般而言，除去其他因素，n_2/n 是关于 α 和 $\{\beta_n\}_{n=1}^{\infty}$ 的函数。不过，在恰当的假设下，ARE 不依赖于这些量。特别地，假设判决 $\widetilde{\delta}_j^{(n)}$ 的形式为

$$\widetilde{\delta}_j^{(n)}(\boldsymbol{y})=\begin{cases}1, & T_j^{(n)}(y_1,\cdots,y_n)>\tau_j^{(n)}\\ \gamma_j, & T_j^{(n)}(y_1,\cdots,y_n)=\tau_j^{(n)}\\ 0, & T_j^{(n)}(y_1,\cdots,y_n)<\tau_j^{(n)}\end{cases}$$

并且假设检验的形式为

$$\begin{aligned}H_0&:Y\sim P_{\theta_0}\\ H_1^{(n)}&:Y\sim P_{\theta_n}\end{aligned}$$

其中 $\theta_n>\theta_0$ 和 $\{P_\theta;\ \theta>\theta_0\}$ 是 Y 的分布族。此时，$H_1^{(n)}$ 收敛到 H_0 的概念可表述为 $\lim\limits_{n\to\infty}\theta_n=\theta_0$。对 $j=0$，1，$n=1$，2，…和 $\theta\geqslant\theta_0$，定义

$$\psi_j^{(n)}(\theta)=E\{T_j^{(n)}(Y_1,\cdots,Y_n)\,|\,Y\sim P_\theta\}$$

和

$$\sigma_j^{(n)}(\theta)=[\mathrm{var}(T_j^{(n)}(Y_1,\cdots,Y_n)\,|\,Y\sim P_\theta)]^{1/2}$$

即 $\psi_j^{(n)}(\theta)$ 和 $\sigma_j^{(n)}(\theta)$ 是 $Y\sim P_\theta$ 时 $T_j(Y)$ 的均值和标准差。

考虑以下正则条件：

1) 存在一个正整数 m，使得 $\psi_j^{(n)}(\theta)$ 的 1 到 $m-1$ 阶导数在 $\theta=\theta_0$ 处等于 0，且

⊖ 具有这个性质的测试序列被认为是连贯的。

$$\frac{\mathrm{d}^m}{\mathrm{d}\theta^m}\psi_j^{(n)}(\theta)\mid\theta=\theta_0>0,\quad j=0,1$$

2）对于 $j=0$，1，存在 $\delta>0$，满足

$$\lim_{n\to\infty}\left[n^{-m\delta}\frac{\mathrm{d}^m}{\mathrm{d}\theta^m}\psi_j^{(n)}(\theta)\bigg|_{\theta=\theta_0}\sigma_j^{(n)}(\theta_0)\right]\triangleq c_j>0 \tag{3.3.24}$$

3）对于 $n=1$，2，…，定义 $\theta_n=\theta_0+Kn^{-\delta}$，有下式成立

$$\lim_{n\to\infty}\left[\frac{\mathrm{d}^m}{\mathrm{d}\theta^m}\psi_j^{(n)}(\theta)\mid_{\theta=\theta_n}\Big/\frac{\mathrm{d}^m}{\mathrm{d}\theta^m}\psi_j^{(n)}(\theta)\mid_{\theta=\theta_0}\right]=1$$

和

$$\lim_{n\to\infty}[\sigma_j^{(n)}(\theta_n)/\sigma_j^{(n)}(\theta_0)]=1$$

4）定义

$$W_j^{(n)}(\boldsymbol{Y})=[T_j^{(n)}(Y_1,\cdots,Y_n)-\psi_j^{(n)}(\theta)]/\sigma_j^{(n)}(\theta_n)$$

则对所有 $w\in\mathbf{R}$ 和 θ 在 $\theta_0\leqslant\theta\leqslant(\theta_0+d)$ 内的均匀分布($d>0$)，有下式成立

$$\lim_{n\to\infty}P_\theta(W_j^{(n)}(Y)\leqslant w)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{w}\mathrm{e}^{-x^2/2}\mathrm{d}x\triangleq\Phi(w)$$

在上述正则性条件约束下，有下面的命题成立。

命题 3.3.1　Pitman-Noether 定理

假设 $\{\widetilde{\delta}_1^{(n)}\}_{n=1}^{\infty}$ 和 $\{\widetilde{\delta}_2^{(n)}\}_{n=1}^{\infty}$ 满足上面四个条件，对于序列 $\theta_n=\theta_0+Kn^{-\delta}$，有

$$\mathrm{ARE}_{1,2}=\eta_1/\eta_2$$

其中 η_j 定义为

$$\eta_j=(c_j)^{1/m\delta},\quad j=0,1$$

c_j 与式(3.3.24)相同。

证明　关于该定理的完整证明可以参见 Noether(1955)，在这里仅给出证明的基本思想。

首先，条件 4)表明 $T_j^{(n)}(Y_1,\cdots,Y_n)$ 在 H_0 假设下可近似为 $N(\psi_j^{(n)}(\theta_0),[\sigma_j^{(n)}(\theta_0)]^2)$。因此，当 $P_\mathrm{F}(\widetilde{\delta}_j)=\alpha$ 时，有

$$\tau_j^{(n)}\sim\sigma_j^{(n)}(\theta_0)\Phi^{-1}(1-\alpha)+\psi_j^{(n)}(\theta_0)$$

类似地，在虚警概率为 α 时，判决 $\widetilde{\delta}_j$ 对应的检测概率可近似表示为

$$\begin{aligned}P_\mathrm{D}(\widetilde{\delta}_j)&\sim1-\Phi\left[\frac{\tau_j^{(n)}-\psi_j^{(n)}(\theta_n)}{\sigma_j^{(n)}(\theta_n)}\right]\\&=1-\Phi\left[\frac{\sigma_j^{(n)}(\theta_0)}{\sigma_j^{(n)}(\theta_n)}\Phi^{-1}(1-\alpha)-\frac{\psi_j^{(n)}(\theta_n)-\psi_j^{(n)}(\theta_0)}{\sigma_j^{(n)}(\theta_n)}\right]\end{aligned}$$

根据条件 1)，有

$$\psi_j^{(n)}(\theta_0)-\psi_j^{(n)}(\theta_0)\sim\left(\frac{\theta_n-\theta_0}{m!}\right)^m\frac{\mathrm{d}^m}{\mathrm{d}\theta^m}\psi_j^{(n)}(\theta)\mid_{\theta=\theta_0}$$

再根据条件 3)，有 $\sigma_j^{(n)}(\theta_0)\sim\sigma_j^{(n)}(\theta_n)$ 成立。代入 $\theta_n=\theta_0+Kn^{-\delta}$ 后，有

$$P_\mathrm{D}(\widetilde{\delta}_j)\sim1-\Phi\left[\Phi^{-1}(1-\alpha)-\frac{K^m}{m!n^{m\delta}}c_j\right] \tag{3.3.25}$$

如果令 $P_\mathrm{D}(\widetilde{\delta}_1)$ 和 $P_\mathrm{D}(\widetilde{\delta}_2)$ 相等，然后分别用 n_1 和 n_2 表示它们各自的样本数，根据

式(3.3.25)可得

$$\frac{c_1}{(n_1)^{m\delta}} \sim \frac{c_2}{(n_2)^{m\delta}}$$

也等价于

$$\frac{n_2}{n_1} \sim \left(\frac{c_2}{c_1}\right)^{1/m\delta} = \frac{\eta_2}{\eta_1} \tag{3.3.26}$$

命题得证。 □

备注

1) η_j 的值被称为测试序列 $\{\widetilde{\delta}_j^{(n)}\}_{n=1}^{\infty}$ 的极限效力。Pitman-Noether 定理表明具有更高效力的序贯判决是更渐近有效的。

2) 大多数信号检测模型都满足正则条件 1)～4)，例如，对边缘密度 $f_\theta(y)$ 相同的独立同分布序列 Y_1，Y_2，…考虑检测统计量为

$$T_j^{(n)}(y_1,\cdots,y_n) = \sum_{k=1}^{n} g_j(y_k), \quad j = 0,1 \tag{3.3.27}$$

则有下式成立

$$\psi_j^{(n)}(\theta) = n\int g_j f_\theta \, \mathrm{d}\mu$$

和

$$\sigma_j^{(n)}(\theta) = \left[n\int (g_j)^2 f_\theta \mathrm{d}\mu - (\psi_j^{(1)}(\theta))^2\right]^{1/2}$$

下面，假设 $\int g_j f_\theta \mathrm{d}\mu = 0$[在 H_0 假设下，$T_j^{(n)}(Y_1,\cdots,Y_n)$ 的均值在任何情况下都不相关，因为任何常数加在 $T_j^{(n)}$ 上等于相同的常数加在判决门限 $\tau_j^{(n)}$ 上]和给定 $m=1$、$\delta=1/2$ 后，$\partial\int g_j f_\theta \mathrm{d}\mu/\partial\theta|_{\theta=\theta_0} > 0$ 是不合理的，因此有

$$\eta_j = \left[\frac{\partial}{\partial\theta}\int g_j f_\theta \mathrm{d}\mu\right]_{\theta=\theta_0}^2 \bigg/ \int g_j^2 f_{\theta_0} \mathrm{d}\mu \tag{3.3.28}$$

因此，假设条件 3)和 4)成立，则能够相对容易地计算出式(3.3.27)所描述的检测器的 ARE。条件 3)是密度函数 f_θ 关于 θ 平滑的条件，同时也是非线性函数 g_j 平滑的条件。如果 $\int g_j^2 f_\theta \mathrm{d}\mu < \infty$ 成立，由中心极限定理[见 Breiman(1968)]，总能得到

$$P_\theta[W_j^{(n)}(Y) \leqslant w] \to \Phi(w)$$

条件 4)要求在$[\theta_0, \theta_0+d]$内一致收敛。不过，就独立同分布样本而言，有下面的结论成立[称为 Berry-Eseen 边界，见 Breiman(1968)]

$$\sup_{w\in\mathbf{R}} |P_\theta[W_j^{(n)}(Y) \leqslant w] - \Phi(w)| \leqslant \frac{4\int |g_j|^3 f_\theta \mathrm{d}\mu}{\sqrt{n}\left[\int g_j^2 f_\theta \mathrm{d}\mu\right]^{3/2}} \tag{3.3.29}$$

因此，对任意正数 a、b 和 d，如果

$$\int |g_j|^3 f_\theta \mathrm{d}\mu \leqslant a \quad 且 \quad \int g_j^2 f_\theta \mathrm{d}\mu \geqslant b \tag{3.3.30}$$

成立，则对于所有 $\theta \in [\theta_0, \theta_0+d]$，均满足条件 4)。事实上，式(3.3.30)成立是非常合

理的。

3) 在备注 2)的检测模型中，可以看到$\{\widetilde{\delta}_j^{(n)}\}_{n=1}^{\infty}$的有效性取决于$\partial\int g_j f_\theta \mathrm{d}\mu/\partial\theta|_{\theta=\theta_0}$的值。这个值可重写成$\partial\int g_j l_\theta f_\theta \mathrm{d}\mu/\partial\theta|_{\theta=\theta_0}$，其中$l_\theta=f_\theta/f_{\theta_0}$。假设$l_\theta$足够光滑，则可以交换积分和微分的次序，进而有

$$\frac{\partial}{\partial\theta}\int g_j f_\theta \mathrm{d}\mu|_{\theta=\theta_0}=\int g_j\left(\frac{\partial}{\partial\theta}l_\theta\right)_{\theta=\theta_0}f_{\theta_0}\mathrm{d}\mu=\int g_j T_{\mathrm{lo}} f_{\theta_0}\mathrm{d}\mu \tag{3.3.31}$$

其中

$$T_{\mathrm{lo}}=\frac{\partial}{\partial\theta}l_\theta|_{\theta=\theta_0} \tag{3.3.32}$$

因此，在上述假设下，式(3.3.28)的效力或有效性将变为

$$\eta_j=\left[\int g_j T_{\mathrm{lo}} f_{\theta_0}\mathrm{d}\mu\right]^2\Big/\int g_j^2 f_{\theta_0}\mathrm{d}\mu \tag{3.3.33}$$

根据式(3.3.33)和施瓦兹不等式[㊀]，对于任意正常数c，式(3.3.27)在$g_j=cT_{\mathrm{lo}}$时取最大值，即形如式(3.3.27)的系统是最有效的。常数c的取值无关紧要，因为c的变化只会导致判决门限的变化。不过，c必须是正数，因为

$$\frac{\partial}{\partial\theta}\int[cT_{\mathrm{lo}}]f_\theta\mathrm{d}\mu\Big|_{\theta=\theta_0}=c\int(T_{\mathrm{lo}})^2 f_{\theta_0}\mathrm{d}\mu>0$$

将$g_j=cT_{\mathrm{lo}}$代入式(3.3.33)，η_j的最大值为

$$\max\eta_j=\int(T_{\mathrm{lo}})^2 f_{\theta_0}\mathrm{d}\mu \tag{3.3.34}$$

这个量将在第 4 章的参数估计中再次出现。

值得注意的是，从第 2 章的观点来看，由非线性操作T_{lo}得到判决器也是判决$\theta=\theta_0$和$\theta>\theta_0$的局部最优判决器。对固定的n，似然比检测器具有最佳性能。因此，对于任意有限的n，任何检测器相对于似然比检测器的效率都不会大于 1。既然基于T_{lo}的检测器是渐近最有效的，因此它渐近等效于似然比检测器。

3.4 序贯检测

在 3.2 节中讨论的都是固定样本数的检测器，即给定样本数设计最优检测器。除此之外，还可以根据检测器的性能来选择样本数。即对观测序列的某些具体实现而言，可能只需几个样本就能做出正确判决；对于另一些实现而言，可能需要更多的样本才能做出正确的判决。这种根据观测序列来选取样本数的检测器通常称为序贯检测器。

假设观测集$\Gamma=\mathbf{R}^\infty$是所有(单边)实序列的集合，且观测值$\{Y_k;\ k=1,2,\cdots\}$满足独立同分布[㊁]。假设检验为

$$\begin{aligned}H_0&:Y_k\sim P_0,\quad k=1,2,\cdots\\H_1&:Y_k\sim P_1,\quad k=1,2,\cdots\end{aligned} \tag{3.4.1}$$

㊀ 施瓦兹不等式：对于常数c，$c\in\mathbf{R}$，当且仅当$f=cg$时，$\left[\int|fg|\mathrm{d}u\right]^2\leqslant\int f^2\mathrm{d}u\int g^2\mathrm{d}u$等号成立。

㊁ 即$\mathbf{R}^\infty=\{y\mid y=\{y_k\}_{k=1}^{\infty}$，其中$y_k\in\mathbf{R}$，$k\geqslant1\}$。

其中，P_0 和 P_1 是$(\mathbf{R}, \mathcal{B})$上的两种分布，$\mathcal{B}$是定义在 $\mathbf{R}$ 上的 Borel σ 代数。序贯判决准则是$(\boldsymbol{\phi}, \boldsymbol{\delta})$的序列对，其中 $\boldsymbol{\phi}=\{\phi_j;\ j=0, 1, 2, \cdots\}$称为终止规则$(\phi_j:\mathbf{R}^j\rightarrow\{0, 1\})$，$\boldsymbol{\delta}=\{\delta_j;\ j=0, 1, 2, \cdots\}$称为终端决策规则，当 $j\geqslant 0$ 时 δ_j 是$(\mathbf{R}^j, \mathcal{B}^j)$上的一个判决准则。

序贯判决准则$(\boldsymbol{\phi}, \boldsymbol{\delta})$的实现步骤如下：对于观测序列$\{y_k;\ k=1, 2, \cdots\}$，准则$(\boldsymbol{\phi}, \boldsymbol{\delta})$做出决定 $\delta_N(y_1, y_2, \cdots, y_N)$，其中 N 是终止时间，定义为 $N=\min\{n\,|\,\phi_n(y_1, y_2, \cdots, y_n)=1\}$。$\boldsymbol{\phi}$确定何时停止采样，即当 $\phi_n(y_1, y_2, \cdots, y_n)=0$ 时，需要增加第 $n+1$ 个采样，当 $\phi_n(y_1, y_2, \cdots, y_n)=1$ 时，停止采样并做出判决。因此，样本数依赖于观测序列，具有随机性。终端决策规则 $\boldsymbol{\delta}$ 决定停止采样后该做出何种判决。例如，固定样本数 n 的决策规则 δ 可用$(\boldsymbol{\phi}, \boldsymbol{\delta})$表示为

$$\begin{aligned} \phi_j(y_1,\cdots,y_j) &= \begin{cases} 0, & j\neq n \\ 1, & j=n \end{cases} \\ \delta_j(y_1,\cdots,y_j) &= \begin{cases} \delta(y_1,\cdots,y_j), & j=n \\ \text{任意值}, & j\neq n \end{cases} \end{aligned} \tag{3.4.2}$$

为了得到式(3.4.1)的最优序贯决策规则，首先考虑该问题的贝叶斯最小风险判决，其中 H_1 和 H_0 假设的先验概率分别为 π_1 和 $\pi_0=1-\pi_1$，C_{ij} 是判决的代价。为了简化处理，假设代价是相同的。当然，其他的代价分配形式同样容易处理。此外，虽然理论上可得到无穷多个独立同分布的观测数据，但从实际出发，观测也是有代价的。如果每次采样的代价 $C>0$，那么 n 个采样的总代价为 nC。

根据上述代价分配方式，给定序贯判决准则的条件风险为

$$R_0(\boldsymbol{\phi},\boldsymbol{\delta}) = E_0\{\delta_N(Y_1,\cdots,Y_N)\} + CE_0\{N\}$$

和

$$R_1(\boldsymbol{\phi},\boldsymbol{\delta}) = 1 - E_1\{\delta_N(Y_1,\cdots,Y_N)\} + CE_1\{N\} \tag{3.4.3}$$

其中，下标与假设对应，N 是随机序列$\{Y_k\}$的停止时间。贝叶斯风险可表示为

$$r(\boldsymbol{\phi},\boldsymbol{\delta}) = (1-\pi_1)R_0(\boldsymbol{\phi},\boldsymbol{\delta}) + \pi_1 R_1(\boldsymbol{\phi},\boldsymbol{\delta}) \tag{3.4.4}$$

最小化 $r(\boldsymbol{\phi}, \boldsymbol{\delta})$的判决准则称为贝叶斯序贯判决准则。

为了方便在贝叶斯框架下理解最优判决准则的结构，首先定义下述函数

$$V^*(\pi_1) \triangleq \min_{\substack{\boldsymbol{\phi},\boldsymbol{\delta} \\ \phi_0=0}} r(\boldsymbol{\phi},\boldsymbol{\delta}), \quad 0\leqslant\pi_1\leqslant 1 \tag{3.4.5}$$

$\phi_0=0$ 意味着在没有观测值时判决不会停止，式(3.4.5)描述了至少一个样本的所有序贯检测的最小贝叶斯风险。可以直接证明 $V^*(\pi)$是关于 π_1 的连续凸函数，满足 $V^*(0)=V^*(1)=C$，如图 3.4.1所示[参见 Ferguson(1967)]。图 3.4.1中同时给出了其他两个判决准则的贝叶斯风险随 π_1 的变化曲线。这两个判决准则均不需要样本，一个选择 H_1(例如，$\phi_0=\delta_0=1$)，另一个选择 H_0(例如，$\phi_0=1-\delta_0=1$)，有

$$r(\boldsymbol{\phi},\boldsymbol{\delta})\,|_{\phi_0=\delta_0=1} = 1-\pi_1$$

$$r(\boldsymbol{\phi},\boldsymbol{\delta})\,|_{\phi_0=1-\delta_0=1} = \pi_1$$

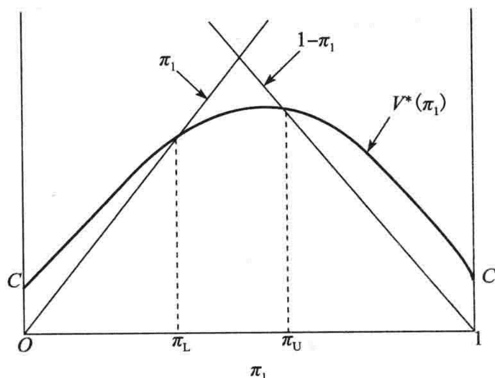


图 3.4.1　均匀误判代价和样本代价为 C 下的贝叶斯序贯准则

注意，这是唯一没有包含在式(3.4.5)最小化公式中的两种代表贝叶斯估计的决策规则[⊖]。在图(3.4.1)中，用 π_U 表示 $r(\boldsymbol{\phi}, \boldsymbol{\delta})|_{\phi_0=\delta_0=1}$ 和 $V^*(\pi_1)$ 的交点对应的横坐标，π_L 表示 $r(\boldsymbol{\phi}, \boldsymbol{\delta})|_{\phi_0=1-\delta_0=1}$ 和 $V^*(\pi_1)$ 的交点对应的横坐标。

结合图 3.4.1 可知：如果 $\pi_1 \leqslant \pi_L$，贝叶斯规则为 $\phi_0=1-\delta_0=1$；如果 $\pi_1 \geqslant \pi_U$，贝叶斯规则为 $\phi_0=\delta_0=1$；如果 $\pi_L<\pi_1<\pi_U$，贝叶斯规则为 $\phi_0=0$ 下的最小化$(\boldsymbol{\phi}, \boldsymbol{\delta})$。所以，对于 $\pi_1 \leqslant \pi_L$ 情况，在选择 H_0 时不需要样本；对于 $\pi_1 \geqslant \pi_L$，在选择 H_1 时也不需要样本；其他情况，则至少需要一个样本。

假设 $\pi_L<\pi_1<\pi_U$，尽管最优判决的形式未定，但它至少需要一个样本。得到一个样本后，如果仍然有无限多的独立同分布样本，则该优化过程与没有样本时是一致的，区别仅在于利用一个样本得到了更多关于哪个假设成立的信息。特别地，用观测数据 Y_1（例如 $\pi_1(y_1)=P(H_1$ 为真 $| Y_1=y_1)$）在 H_1 下的后验 $\pi_1(y_1)$代替先验概率 π_1 后，能得到与图 3.4.1 类似的结果。不过，横坐标变量 $\pi_1(y_1)$要替换为 π_1。此外，由于样本独立，任何有关 Y_1 的信息不会影响 V^* 的形状。综合以上分析，在得到一个样本后，如果 $\pi_1(y_1) \leqslant \pi_L$，则判决为 H_0；如果 $\pi_1(y_1) \geqslant \pi_U$，则判决为 H_1；如果 $\pi_L<\pi_1(y_1)<\pi_U$，则继续增加样本。

如果 $\pi_L<\pi_1<\pi_U$ 且 $\pi_L<\pi_1(y_1)<\pi_U$，则最优判决至少需要两个样本。此时，可采用相同的方法根据新的先验值 $\pi_1(y_1, y_2)=P(H_1$ 为真 $| Y_1=y_1, Y_2=y_2)$进行重新判决。将此推广到任意样本数，可以看到贝叶斯序贯检测会连续采样，直到 $\pi_1(y_1, \cdots, y_n) \triangleq P(H_1$ 为真 $| Y_1=y_1, \cdots, Y_n=y_n)$落在区间$(\pi_L, \pi_U)$之外，并根据 $\pi_1(y_1, \cdots, y_n) \leqslant \pi_L$ 选择 H_0，$\pi_1(y_1, \cdots, y_n) \geqslant \pi_U$ 选择 H_1。在 $n=0$ 时，用 $\pi_1(y_1, \cdots, y_n)$表示 π_1 后，贝叶斯序贯检测的停止准则为

$$\phi_n(y_1,\cdots,y_n)=\begin{cases}0, & \pi_L<\pi_1(y_1,\cdots,y_n)<\pi_U \\ 1, & \text{其他}\end{cases} \tag{3.4.6}$$

判决准则为

$$\delta_n(y_1,\cdots,y_n)=\begin{cases}1, & \pi_1(y_1,\cdots,y_n) \geqslant \pi_U \\ 0, & \pi_1(y_1,\cdots,y_n) \leqslant \pi_L\end{cases} \tag{3.4.7}$$

图 3.4.2 描述了贝叶斯的判决过程。$\pi_1(Y_1, \cdots, Y_n)$在 H_1 假设下收敛到 1，在 H_0 假设下收敛到 0，判决以概率 1 终止。最优判决只需已知 π_L 和 π_U，以及计算 $\pi_1(y_1, \cdots, y_n)$。不幸的是，除了一些特殊情况(例如 Y_n 只取离散值)，一般很难获得 π_L 和 π_U。相反，计算后验概率 $\pi_1(y_1, \cdots, y_n)$相对容易。特别地，假设 P_0 和 P_1 的密度为 p_0 和 p_1，则根据贝叶斯公式，$\pi_1(y_1, \cdots, y_n)$可以写为

$$\pi_1(y_1,\cdots,y_n)=\frac{\pi_1\prod_{k=1}^{n}p_1(y_k)}{\pi_0\prod_{k=1}^{n}p_0(y_n)+\pi_1\prod_{k=1}^{n}p_1(y_n)}=\frac{\pi_1\lambda_n(y_1,\cdots,y_n)}{\pi_0+\pi_1\lambda_n(y_1,\cdots,y_n)} \tag{3.4.8}$$

其中 λ_n 为 n 个样本的似然比，表示为

⊖ 可以看到当 $\phi_0=1$ 时随机化是无效的，因为如果当概率为 γ 时选 $\delta_0=1$，当概率为 $1-\gamma$ 时选 $\delta_0=0$，则贝叶斯风险 $\gamma(1-\pi_1)+(1-\gamma)\pi$ 永远大于 $\min\{\pi_1, 1-\pi_1\}$。

$$\lambda_n(y_1,\cdots,y_n)=\prod_{k=1}^{n}[p_1(y_n)/p_0(y_n)] \tag{3.4.9}$$

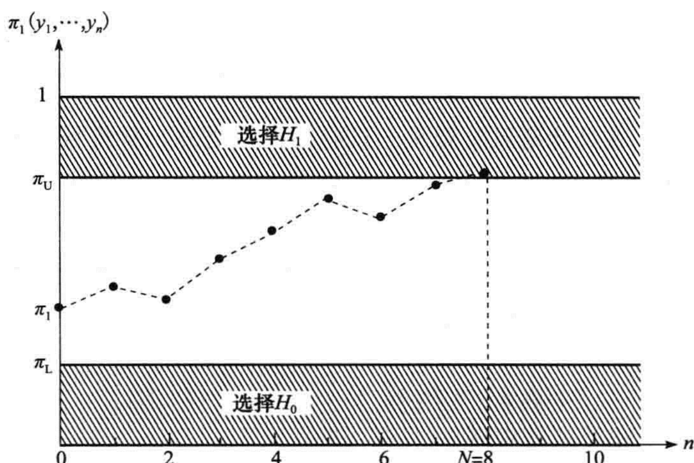


图 3.4.2　贝叶斯序贯检测的实现

从表示的一致性出发，定义 $\lambda_0=1$。因为式(3.4.8)随 λ_n 单调递增，则式(3.4.6)和式(3.4.7)可重写成

$$\phi_n(y_1,\cdots,y_n)=\begin{cases}0, & \underline{\pi}<\lambda_n(y_1,\cdots,y_n)<\overline{\pi}\\ 1, & \text{其他}\end{cases} \tag{3.4.10}$$

和

$$\delta_n(y_1,\cdots,y_n)=\begin{cases}1, & \lambda_n(y_1,\cdots,y_n)\geqslant\overline{\pi}\\ 0, & \lambda_n(y_1,\cdots,y_n)\leqslant\underline{\pi}\end{cases} \tag{3.4.11}$$

其中，$\underline{\pi}\triangleq\pi_0\pi_L/\pi_1(1-\pi_L)$，$\overline{\pi}\triangleq\pi_0\pi_U/\pi_1(1-\pi_U)$。贝叶斯序贯检测会连续采样，直至似然比落在间隔$(\underline{\pi},\ \overline{\pi})$之外，然后根据 λ_n 是否小于 $\underline{\pi}$ 或大于 $\overline{\pi}$ 来决定是 H_0 或 H_1。

式(3.4.10)和式(3.4.11)是序贯概率比检测(SPRT)的特例。对满足 $0<A\leqslant1\leqslant B<\infty$ 的任意实数 A 和 B，用 A 和 B 分别代替 $\underline{\pi}$ 和 $\overline{\pi}$ 后，以 A 和 B 为边界的 SRPT(记为 SPRT(A，B))可以由式(3.4.10)和式(3.4.11)定义，且在 $A=B$ 时，可任意判决。显然，SPRT(A，B)会持续采样，直到似然比 λ_n 落在边界 A 和 B 之外，然后根据 $\lambda_n\geqslant B$ 判决 H_1，根据 $\lambda_n\leqslant A$ 判决 H_0。如果 $A=1<B$，则判决为 H_0；如果 $A<B=1$，则判决为 H_1；如果 $A=B=1$，则可任意判决。

例 3.4.1　直流信号的序贯检测

考虑在加性独立同分布噪声中检测直流信号的问题，即

$$\begin{aligned}H_0: Y_k&=N_k, & k&=1,2,\cdots\\ H_1: Y_k&=N_k+\theta, & k&=1,2,\cdots\end{aligned} \tag{3.4.12}$$

其中 $\theta>0$，$\{N_k\}_{k=1}^{\infty}$是服从 $N(0,\ \sigma^2)$分布的独立同分布噪声样本序列。样本数为 n 的似然比可以表示为

$$\lambda_n(y_1,\cdots,y_n)=\exp\left\{\frac{\theta}{\sigma^2}\sum_{k=1}^{n}(y_k-\theta/2)\right\} \tag{3.4.13}$$

只要似然比落在 A 和 B 之间，SPRT(A, B)就会继续采样。取对数后，SPRT(A, B)的每一步均计算 $\frac{\theta}{\sigma^2}\sum_{k=1}^{n}\left(y_k-\frac{\theta}{2}\right)$，并与 $\log A$ 和 $\log B$ 进行比较。如果 $\frac{\theta}{\sigma^2}\sum_{k=1}^{n}\left(y_k-\frac{\theta}{2}\right)$ 大于 $\log B$，则停止采样并选择 H_1，如果 $\frac{\theta}{\sigma^2}\sum_{k=1}^{n}\left(y_k-\frac{\theta}{2}\right)$小于 $\log A$，也停止采样并选择 H_0。

SPRT($\underline{\pi}$, $\overline{\pi}$)除了在贝叶斯风险意义上具有最优性之外，从信号检测的角度出发，还具有另一种最优特性。在 Wald-Wolfowitz 定理(类似于尼曼-皮尔逊引理)中对这种最优特性进行了总结。

对序贯判决准则($\boldsymbol{\phi}$, $\boldsymbol{\delta}$)，分别用 $P_{\mathrm{F}}(\boldsymbol{\phi}, \boldsymbol{\delta})$和 $P_{\mathrm{M}}(\boldsymbol{\phi}, \boldsymbol{\delta})$表示虚警概率和漏检概率，即

$$P_{\mathrm{F}}(\boldsymbol{\phi},\boldsymbol{\delta}) = P[\delta_N(Y_1,\cdots,Y_N) = 1 \mid H_0]$$

和

$$P_{\mathrm{M}}(\boldsymbol{\phi},\boldsymbol{\delta}) = P[\delta_N(Y_1,\cdots,Y_N) = 0 \mid H_1]$$

同时用 $N(\boldsymbol{\phi})$表示与 $\boldsymbol{\phi}$ 相关的随机停止时间，即

$$N(\boldsymbol{\phi}) = \min\{n \mid \phi_n(Y_1,\cdots,Y_n) = 1\}$$

$N(\boldsymbol{\phi})$也称为 $\boldsymbol{\phi}$ 的样本数。那么有下面的结论。 ◀

命题 3.4.1 Wald-Wolfowitz 定理

假设($\boldsymbol{\phi}_0$, $\boldsymbol{\delta}_0$)是 SPRT(A, B)的序贯判决准则，对于满足 $P_{\mathrm{F}}(\boldsymbol{\phi}, \boldsymbol{\delta})\leqslant P_{\mathrm{F}}(\boldsymbol{\phi}_0, \boldsymbol{\delta}_0)$和 $P_{\mathrm{M}}(\boldsymbol{\phi}, \boldsymbol{\delta})\leqslant P_{\mathrm{M}}(\boldsymbol{\phi}_0, \boldsymbol{\delta}_0)$的任意序贯判决准则($\boldsymbol{\phi}$, $\boldsymbol{\delta}$)而言，有下式成立

$$E\{N(\boldsymbol{\phi}) \mid H_j\} \geqslant E\{N(\boldsymbol{\phi}_0) \mid H_j\}, \quad j = 0,1$$

显然，给定虚警概率和漏检概率后，在所有的序贯判决准则中，SPRT 准则需要的期望样本数是最少的。又因为固定样本数检测是序贯检测的特例，所以在同等性能条件下，SPRT 使用的平均样本数小于或等于固定样本数检测的样本数。事实上，当给定期望样本数时，没有一个序贯判决准则有比 SPRT 更小的误差概率[见 Ferguson(1967)]。Wald-Wolfowitz 定理的有效性可根据 SPRT 的贝叶斯最优性进行证明，具体证明过程从略。

由 Wald-Wolfowitz 定理可知，给定 P_{F} 和 P_{M} 后，SPRT 是平均样本数最少的检测器。即通过增大样本数后，SPRT(A, B)可实现任意的 P_{F} 和 P_{M}。与 SPRT(A, B)有关的另一个问题是如何选择 A 和 B 来达到所期望的性能。下面的分析能部分回答这个问题。

假设($\boldsymbol{\phi}$, $\boldsymbol{\delta}$)是 SPRT(A, B)的判决准则且有 $A<1<B$，令 $\alpha=P_{\mathrm{F}}(\boldsymbol{\phi}, \boldsymbol{\delta})$，$\gamma=(1-\beta)=P_{\mathrm{M}}(\boldsymbol{\phi}, \boldsymbol{\delta})$和 $N=N(\phi)$。注意，($\boldsymbol{\phi}$, $\boldsymbol{\delta}$)的拒绝区域可写成

$$\Gamma_1 = \{y \in \mathbf{R}^\infty \mid \lambda_N(y_1,\cdots,y_N) \geqslant B\} = U_{n=1}^{\infty} Q_n$$

其中 $Q_n=\{y\in\mathbf{R}^\infty \mid N=n \text{ 且 } \lambda_n(y_1, y_2, \cdots, y_n)\geqslant B\}$。考虑 $m\neq n$、Q_n 和 Q_m 互斥后，有

$$\alpha = P(\lambda_N(Y_1,\cdots,Y_N) \geqslant B \mid H_0) = \sum_{n=1}^{\infty}\int_{Q_n}\prod_{k=1}^{n}[p_0(y_k)u(\mathrm{d}y_k)]$$

对于 Q_n，有 $\prod_{k=1}^{n} p_0(y_k) \leqslant B^{-1}\prod_{k=1}^{n} p_1(y_k)$，所以

$$\alpha \leqslant B^{-1}\sum_{k=1}^{n}\int_{Q_n}\prod_{k=1}^{n}[p_1(y_k)u(\mathrm{d}y_k)] = B^{-1}P(\lambda_N(Y_1,\cdots,Y_N) \geqslant B \mid H_1) = B^{-1}(1-\gamma)$$

类似地

$$\gamma = P(\lambda_N(Y_1,\cdots,Y_N) \leqslant A \mid H_1) \leqslant AP(\lambda_N(Y_1,\cdots,Y_N) \leqslant A \mid H_0) = A(1-\alpha)$$

综上可得

$$B \leqslant (1-\gamma)/\alpha \quad 和 \quad A \geqslant \gamma/(1-\alpha) \tag{3.4.14}$$

假设似然比过边界后的量[$\lambda_N(Y_1,\cdots,Y_N)-B$和$A-\lambda_N(Y_1,\cdots,Y_N)$]可忽略，根据给定的$\alpha$和$\gamma$，由不等式(3.4.14)可近似确定边界$A$和$B$。如果$N$远大于平均样本值，那么这种近似是非常精确的。因此，在假设$\lambda_N(Y_1,\cdots,Y_N)\approx A$和$\lambda_N(Y_1,\cdots,Y_N)\approx B$后，不等式(3.4.14)可近似写成等式，即

$$B \approx (1-\gamma)/\alpha \quad 和 \quad A \approx \gamma/(1-\alpha) \tag{3.4.15}$$

这些近似也称为Wald近似。

假设α_d和γ_d是设计的误差概率，则可根据式(3.4.15)来选择实际边界，即

$$A_a = \gamma_d/(1-\alpha_d) \quad 和 \quad B_a = (1-\gamma_d)/\alpha_d \tag{3.4.16}$$

对应的实际误差概率α_a和γ_a满足不等式(3.4.14)，因此有下式成立

$$\alpha_a/(1-\gamma_a) \leqslant B_a^{-1} = \alpha_d/(1-\gamma_d)$$

和

$$\gamma_a/(1-\alpha_a) \leqslant A_a = \gamma_d/(1-\alpha_d)$$

进而可得到

$$\alpha_a \leqslant \alpha_d(1-\gamma_a)/(1-\gamma_d) \leqslant \alpha_d/(1-\gamma_d)$$

和

$$\gamma_a \leqslant \gamma_d(1-\alpha_a/(1-\alpha_d) \leqslant \gamma_d/(1-\alpha_d) \tag{3.4.17}$$

例如，如果$\gamma_d=\alpha_d$，有

$$\alpha_a \leqslant \alpha_d + O(\alpha_d^2)$$

和

$$\gamma_a \leqslant \gamma_d + O(\gamma_d^2)$$

值得注意的是：选择合适的边界A和B后，式(3.4.17)能够确保获得任意好的误差概率性能。此外，不等式(3.4.17)严格成立且不依赖过边界量可忽略的近似[当然，在式(3.3.15)中需要利用该近似]。上述分析的特别之处在于分析过程与Y_k的真实分布无关，这正是SPRT相对于固定样本检测的优势。

虽然SPRT可以任意减小误差概率，但这需要以增加训练样本为代价。假设α和γ足够小，满足$A\approx\gamma$及$B\approx 1/\alpha$，那么当γ和α继续减小时，A将变小而B将变大。即区间(A, B)将扩大，给定观测后，SPRT(A, B)需要更多的样本。根据前面的分析，控制SPRT的误差概率可不用考虑观测值的分布。不过，用来衡量期望的训练样本数与数据分布密切相关，同时它也是用于比较各种序列检测器性能的主要指标。因此，如何估计序列检测器的期望样本数就显得非常重要。为了便于给出具体计算方法，首先考虑更一般的序贯检测如下。

对任意$a<0<b$和函数$g:\mathbf{R}\rightarrow\mathbf{R}$，由规则$(\boldsymbol{\phi}, \boldsymbol{\delta})$定义的序贯判决准则为ST$(a, b; g)$，其中$(\boldsymbol{\phi}, \boldsymbol{\delta})$可表示为

$$\phi_j(y_1,\cdots,y_j)=\begin{cases}0, & a<\sum_{k=1}^{j}g(y_k)<b\\1, & 其他\end{cases}$$

和

$$\delta_j(y_1,\cdots,y_j)=\begin{cases}1, & \sum_{k=1}^{j}g(y_k)\geqslant b\\0, & \sum_{k=1}^{j}g(y_k)\leqslant a\end{cases}$$

事实上，对 $0<A<1<B<\infty$ 和 $p_1/p_0<\infty$，进行 $a=\log A$，$b=\log B$ 和 $g=\log(p_1/p_0)$代换后，SPRT(A，B)等价于 ST(a，b；g)。对于例 3.4.1 的情况，选择 $g(x)=\theta(x-\theta/2)/\sigma^2$，则 SPRT（$A$，$B$）等价于 ST（$a$，$b$；$g$）。图 3.4.3为 ST($a$，$b$；$g$)的判决示意图。

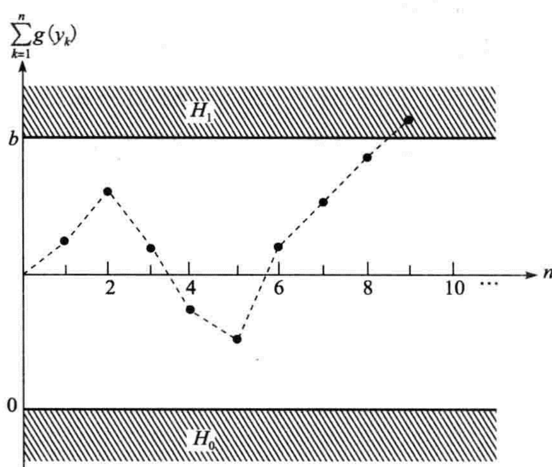


图 3.4.3 ST(a，b；g)判决示意

在 ST（a，b；g）框架下，根据 Wald [见 Ferguson(1967)]有下面的结论成立。

命题 3.4.2 序贯分析的基本恒等式

假设($\boldsymbol{\phi}$，$\boldsymbol{\delta}$)＝ST（a，b；g），同时定义 $N=\min\{n \mid \phi_n(Y_1,\cdots,Y_n)=1\}$ 和 $S_n=\sum_{k=1}^{n}g(Y_k)$，令 M_j 表示 H_j 假设下随机变量 $g(Y_1)$的矩母函数，即

$$M_j(t)=E\{\exp\{tg(Y_1)\}\mid H_j\},\quad j=0,1$$

如果 $P(g(Y_1)=0 \mid H_j)\neq 1$ 且 $P(|g(Y_1)|<\infty \mid H_j)=1$，对使 $M_j(t)<\infty$成立的任意实数，有

$$E\{\exp\{tS_N\}[M_j(t)]^{-N}\mid H_j\}=1$$

该等式也称为 Wald 恒等式。它可由随机采样定理中的马丁格尔策略直接得到[参见 Breiman(1968)]，这里不给予证明。命题 3.4.2 的结论如下。

命题 3.4.3 Wald 恒等式的推论

在命题 3.4.2 的假设下，在 $t=0$ 的邻域内假设 $M_j<\infty$。定义 $\mu_j=E\{g(Y_1)\mid H_j\}$和 $\sigma_j^2=\mathrm{var}(g(Y_1)\mid H_j)$后，有

$$(a)\ E\{S_N\mid H_j\}=\mu_jE\{N\mid H_j\}$$

和

$$(b)\ E\{(S_N-N\mu_j)^2\mid H_j\}=\sigma_j^2E\{N\mid H_j\}$$

利用 Wald 恒等式及其推论，可将 Wald 近似式(3.3.15)扩展到 ST（a，b；g）判决。具体如下。

假设可以找到两个非零数 t_0 和 t_1，在 $j=0$ 和 1 时满足 $M_j(t_j)=1$[它们的存在性由

$P_j(g(Y_1)<0)>0$、$P_j(g(Y_1)>0)>0$ 和 $\mu_j\neq 0$ 保证，参见 Ferguson(1967)]。那么 Wald 恒等式意味着

$$E\{\exp\{t_j S_N\}\mid H_j\}=1,\quad j=0,1 \tag{3.4.18}$$

如果忽略"超过边界的值"，在 H_0 假设下，S_N 是一个离散随机变量，分别以概率 $P_F(\boldsymbol{\phi},\boldsymbol{\delta})\triangleq\alpha$ 和 $1-\alpha$；在 H_1 假设下，S_N 分别以概率 $\gamma=P_M(\boldsymbol{\phi},\boldsymbol{\delta})$ 和 $1-\gamma$ 取 a 和 b。利用式(3.4.18)，可得到

$$(1-\alpha)e^{t_0 a}+\alpha e^{t_0 b}\approx 1$$

和

$$\gamma e^{t_1 a}+(1-\gamma)e^{t_1 b}\approx 1$$

进而有

$$\alpha\approx(1-e^{t_0 a})/(e^{t_0 b}-e^{t_0 a}) \tag{3.4.19}$$

和

$$\gamma\approx(1-e^{t_1 b})/(e^{t_1 a}-e^{t_1 b}) \tag{3.4.20}$$

作为示例，选取 $g=\log(p_1/p_0)$，则 $\mathrm{ST}(a,b;g)=\mathrm{SPRT}(e^a,e^b)$，可得

$$M_0(t)=\int_{\mathbf{R}}\exp\{t\log[p_1(y)/p_0(y)]\}p_0(y)\mu(dy)=\int_{\mathbf{R}}[p_1(y)]^t[p_0(y)]^{1-t}\mu(dy)$$

注意

$$M_0(1)=\int_{\mathbf{R}}p_1(y)\mu(dy)=1$$

得到 $t_0=1$。类似地，根据

$$M_1(t)=\int_{\mathbf{R}}[p_1(y)]^{t+1}[p_0(y)]^{-t}\mu(dy)$$

可得 $t_1=-1$。在定义 $A=e^a$ 和 $B=e^b$ 之后，式(3.4.19)和式(3.4.20)可以写成

$$\alpha\approx(1-A)/(1-B)\quad 和\quad \gamma\approx(1-B^{-1})/(A^{-1}-B^{-1})$$

它们等价于式(3.4.15)的 Wald 近似。

如果 $\mu_0=0$，则不存在使 $M_0(t_0)=1$ 的非零数 t_0。但在这种情况下，Wald 恒等式的推论隐含了 $E\{S_N\mid H_0\}=0$，进而有

$$a(1-\alpha)+b\alpha\approx 0$$

或

$$\alpha\approx -a/(b-a) \tag{3.4.21}$$

类似地，如果 $\mu_1=0$，有

$$\gamma\approx -b/(a-b) \tag{3.4.22}$$

采用与示例中相同的分析方法，可以根据命题 3.4.3 来近似 $\mathrm{ST}(a,b;g)$ 的预期样本数。首先，假设 $\mu_0\neq 0$，然后忽略超过边界的量，利用推论，有

$$E\{N\mid H_0\}=\frac{1}{\mu_0}E\{S_N\mid H_0\}\approx\frac{1}{\mu_0}(a(1-\alpha)+b\alpha)$$

如果 $\mu_0=0$，则需要利用

$$E\{N\mid H_0\}=\frac{1}{\sigma_0^2}E\{S^2\mid H_0\}\approx\frac{1}{\sigma_0^2}(a^2(1-\alpha)+B^2\alpha)$$

类似的表达式对 $E\{N \mid H_1\}$ 也成立。结合式(3.4.19)和式(3.4.20)或者式(3.4.21)和式(3.4.22)的都可以消去(a, b)和(α, γ)。例如，如果 $\mu_j=0$，有

$$E\{N|H_j\} \approx -ab/\sigma_j^2$$

对于 $\mu_j \neq 0$ 的 SPRT(e^a，e^b)问题，利用 Wald 近似后，有

$$\begin{aligned} E\{N|H_0\} &\approx \frac{1}{\mu_0}\left[(1-\alpha)\log\frac{\gamma}{1-\alpha}+\alpha\log\frac{1-\gamma}{\alpha}\right] \\ E\{N|H_1\} &\approx \frac{1}{\mu_1}\left[\gamma\log\frac{\gamma}{1-\alpha}+(1-\gamma)\log\frac{1-\gamma}{\alpha}\right] \end{aligned} \tag{3.4.23}$$

现在考虑下面的例子。

例 3.4.2 序贯检测与固定样本集检测对比

为了说明序贯检测(SPRT)能比固定样本集检测(FSS)具有更少的样本数，再次考虑例 3.4.1 中的直流信号检测问题。

最优 FSS 检测器是似然比检测器，根据式(3.2.31)，似然比检测器的误差概率 α 和相关的 γ 与样本数的关系为

$$1-\gamma = 1-\Phi(\Phi^{-1}(1-\alpha)-n^{1/2}\theta/\sigma) \tag{3.4.24}$$

在给定 α 和 γ 后，FSS 的似然比检测器所需样本数可以通过对式(3.3.24)求逆得到，即

$$n_{\text{FSS}} = \left\lceil \sigma^2[\Phi^{-1}(1-\alpha)-\Phi^{-1}(\gamma)]^2/\theta^2 \right\rceil \tag{3.4.25}$$

其中$\lceil x \rceil$表示不小于 x 的最小整数。

误差概率选择边界后，SPRT 的期望样本数可以根据式(3.4.23)得到。求解式(3.4.23)需要计算 μ_0 和 μ_1，即

$$\begin{aligned} \mu_j &= E\{\log[p_1(Y_1)/p_0(Y_1)]|H_j\} = \frac{\theta}{\sigma^2}E\{(Y_1-\theta/2)|H_j\} \\ &= \begin{cases} -\theta^2/2\sigma^2, & j=0 \\ +\theta^2/2\sigma^2, & j=1 \end{cases} \end{aligned} \tag{3.4.26}$$

为了简化，假设 $\alpha=\gamma$ 后，有

$$E\{N|H_1\} = E\{N|H_0\} \approx 2\sigma^2\left[\gamma\log\frac{\gamma}{1-\alpha}+(1-\gamma)\log\frac{1-\gamma}{\alpha}\right] \tag{3.4.27}$$

当 $\alpha=\gamma=0.1$ 和 $\theta^2/\sigma^2=1$ 时，计算式(3.4.25)和式(3.4.27)得到 $n_{\text{FSS}}\approx 22$ 和 $E\{N \mid H_j\}\approx 9$。因此，在这种情况下，SPRT 平均使用的样本数是 FSS 的一半。结合式(3.4.25)和式(3.4.27)，有

$$\lim_{\alpha=\gamma\to 0}\frac{E\{N|H_j\}}{n_{\text{FSS}}} = 1/4 \tag{3.4.28}$$

对于足够小的误差概率($\alpha=\gamma$)而言，SPRT 平均所需要的样本数为 FSS 的 1/4。 ◀

例 3.4.2 说明，约束检测性能后，SPRT 做出判决所需的平均样本数少于最优 FSS。这在需要进行大量相同判决时非常有效。例如在搜索雷达中，需要对搜索区域内的每个检测单元进行判决来确定是否存在目标。此时，采用 SPRT 将比 FSS 具有更短的判决时间。

既然 SPRT 能获得性能增益，那么为什么不用 SPRT 代替 FSS 呢？遗憾的是，SPRT 在实际应用中存在之前的分析中没有涉及的几个缺点。其中一个缺点就是 SPRT 的样本数

以概率1有限，而不是有界。在观测数据清晰可判时，SPRT 通过快速判决来降低样本数；不过，如果观测数据的模糊性较强，则 SPRT 需要大量的样本。虽然 Wald-Wolfowitz 定理表明对这两种方式进行平均是有利的，但是，对很多应用来说，即使是偶尔发生的长时判决过程也是不允许的。对于这一问题，可以通过修改 SPRT 的停止准则来解决，即在样本数超过最大值后进行硬判决。这种类型的检测也称为截断 SPRT，且如果截断点不是太小，也可具备 SPRT 的良好性能。

SPRT 的另一个缺点是需要精确知道 p_0 和 p_1。例如，在例 3.4.1 的直流信号检测问题中，进行序贯最大似然判决必须知道信号值 θ，这与具有一致最大势的固定样本集尼曼-皮尔逊检测器形成鲜明对比。例如，当真实的位置参数只是假设的一半时，H_1 假设下的统计判决量 $\theta\sum_{k=1}^{n}(y_k-\theta/2)/\sigma^2$ 将在零附近波动，进而增加了判决时间。尽管通过截断处理能够在一定程度上缓解此类问题，但这确实是限制序贯检测获得应用的主要原因。

序贯检测的第三个缺点是样本不满足独立同分布时，原始的 SPRT 作为最优贝叶斯检测就会失效。这是因为如果之前和之后的样本不独立，且每一步的样本分布特性不相同，那么基于贝叶斯最优判决的 SPRT 推导过程将不再成立。

除去以上三个缺点之外，SPRT 的诸多优点使其得到了广泛应用。在 Tantaratana (1986)的综述性论文中给出了 SPRT 的其他性质和具体修正模型。

3.5　非参数检测和稳健检测

在第2章和第3.2节中，研究了观测数据的统计特性服从某种分布下的假设检验和信号检测问题。然而，它们的共同前提是观测数据的概率分布已知，或者仅仅是部分参数未知。通常在实际情况下，明确已知观测数据的分布特性是不切实际的，有时候甚至不能假定它们近似已知。在这种情况下，之前介绍的检测方法就不再有效。为此，必须设计观测数据的分布特性未知下的检测准则。在这个背景下，非参数检测和稳健检测是两个可供选择的设计原则。一般而言，非参数化方法解决了只知道噪声的大致统计特性下的信号检测问题，稳健检测方法用于噪声统计特性非精确已知的情况。首先，在这里简要描述一下这两种检测方法。

考虑下列一般的二元复合假设检验问题，其观测序列服从独立同分布

$$\begin{aligned} H_0: Y_k \sim P \in \mathcal{P}_0, \quad k = 1,2,\cdots,n \\ H_1: Y_k \sim P \in \mathcal{P}_1, \quad k = 1,2,\cdots,n \end{aligned} \tag{3.5.1}$$

其中$\mathcal{P}_0$ 和$\mathcal{P}_1$ 是对于观测数据的两类不完全重叠的边缘分布。如果$\mathcal{P}_0$ 和$\mathcal{P}_1$ 可以用一个实数或者向量参数表示，那么就可看作是参数化假设检验问题。例如，在第 2.5 节中讨论的复合假设检验问题就是参数化假设检验问题。如果$\mathcal{P}_0$ 和$\mathcal{P}_1$ 不能参数化表示，那么式(3.5.1)就是非参数化假设检验问题。在非参数化假设检验中，普遍认为$\mathcal{P}_0$ 和$\mathcal{P}_1$ 的分布太宽而不能用有限维参数表示。

非参数假设检验问题的一个例子就是位置检验，即

$$\begin{aligned} H_0: Y_k = N_k, \qquad k = 1,2,\cdots,n \\ H_0: Y_k = N_k + \theta, \quad k = 1,2,\cdots,n \end{aligned} \tag{3.5.2}$$

其中$\{N_k\}_{k=1}^{n}$是一个仅知其边缘分布关于零对称的独立同分布序列。一般来说，这种模型见于噪声环境下恒定信号的检测问题中，检测序列除了同分布(同一性)、独立和对称性外，其他统计学特征完全未知。由于所有对称分布类一定不是有限维的，所以是典型的非参数问题。

稳健和非参数假设检验都属于非参数化假设理论框架。下面章节首先从非参数检验开始讨论，然后讨论稳健检测问题。

3.5.1 非参数检测

一般来说，设计非参数检测的目的是希望应用于广义的$\mathcal{P}_0$和$\mathcal{P}_1$上，且具有某些不变特性。通常这些检测是简单的，它仅用观测数据的一些简单特性(比如符号、秩等)来代替数据值。总之，非参数检测的性能特征中虚警概率几乎总是保持不变。因此，对式(3.5.1)中的非参数检验器的一个标准定义就是：在$\mathcal{P}_0$下虚警概率为常数。在Y由观测序列Y_1，…，Y_n组成的条件下考虑序列检验问题，即基于$\{\delta_n(y_1, \cdots, y_n)\}_{n=1}^{\infty}$的渐近非参数检测。也就是说，对于所有的$P\in\mathcal{P}_0$，$\lim\limits_{n\to\infty}P_F(\delta_n)$为常数。

非参数检验和检测器在雷达和声呐中得到了广泛的应用。在这些应用中，非参数检测器有时候也称为恒定虚警概率(CFAR)检测器。下面给出一些常用的非参数方法。

1. 符号检测

假设实观测序列Y_1，…，Y_n满足独立同分布。定义参数p满足$p=P(Y_1>0)$且考虑如下假设检验

$$\begin{aligned} &H_0: p = \frac{1}{2} \\ &H_1: \frac{1}{2} < p < 1 \end{aligned} \tag{3.5.3}$$

简言之，式(3.5.3)是零中值相对于非零中值的假设检验问题。需要指出的是，这个模型出现在许多应用当中，例如在式(3.5.2)中，$\theta>0$就是该模型的子集。

式(3.5.3)是典型的非参数的假设检验问题。根据式(3.5.1)，有

$$\mathcal{P}_0 = \left\{P\in\mathcal{M} \,|\, P(0,\infty) = \frac{1}{2}\right\} \tag{3.5.4}$$

$$\mathcal{P}_1 = \left(P\in\mathcal{M} \,|\, 1 > P(0,\infty) > \frac{1}{2}\right) \tag{3.5.5}$$

这里$\mathcal{M}$表示定义在$(\mathbf{R}, \mathcal{B})$上的分布类且都不能由有限维参量表示。

为了推导式(3.5.3)的最优判决准则，首先在$\mathcal{P}_1$中选择某个任意分布的Q_1。为方便说明，假设任意分布Q_1的概率密度函数为q_1(当然，没有这样的假设同样可完成下面的推导)。定义两个函数

$$q_1^{+}(x) = \begin{cases} q_1(x), & x > 0 \\ 0, & x \leqslant 0 \end{cases}$$

和

$$q_1^{-}(x) = \begin{cases} 0 & x > 0 \\ q_1(x) & x \leqslant 0 \end{cases} \tag{3.5.6}$$

在$(\mathbf{R},\mathcal{B})$上定义概率密度函数q_0为

$$q_0(x)=\frac{q_1^+(x)}{2\int_0^{\infty}q_1(t)\mathrm{d}t}+\frac{q_1^-(x)}{2\int_{-\infty}^{0}q_1(t)\mathrm{d}t} \tag{3.5.7}$$

注意，与概率密度函数q_0所对应的任意分布Q_0属于假设检验$\mathcal{P}_0$的范畴，这是因为

$$Q_0(0,\infty)=\int_0^{\infty}q_0(x)\mathrm{d}x=\frac{\int_0^{\infty}q_1^+(x)\mathrm{d}x}{2\int_0^{\infty}q_1(x)\mathrm{d}x}=\frac{1}{2} \tag{3.5.8}$$

考虑如下简单的假设对

$$\begin{aligned}H_0':Y_k&\sim Q_0,\quad k=1,\cdots,n\\ H_1':Y_k&\sim Q_1,\quad k=i,\cdots,n\end{aligned} \tag{3.5.9}$$

根据尼曼-皮尔逊准则，似然比检验是虚警概率为α的最大势检验，似然比按照下式计算

$$L(\mathbf{y})=\prod_{k=1}^{n}\frac{q_1(y_k)}{q_0(y_k)} \tag{3.5.10}$$

注意

$$\frac{q_1(y_k)}{q_0(y_k)}=\begin{cases}2Q_1^+, & y_k>0\\ 2(1-Q_1^+), & y_k\leqslant 0\end{cases} \tag{3.5.11}$$

其中$Q_1^+\triangleq Q_1((0,\infty))$，因此$L(y)$可以写为

$$L(\mathbf{y})=2^n[Q_1^+]^n[Q_1^+/(1-Q_1^+)]^{t(\mathbf{y})} \tag{3.5.12}$$

其中

$$t(\mathbf{y})\triangleq\sum_{k=1}^{n}u(y_k) \tag{3.5.13}$$

u表示单位阶跃函数，定义为$u(x)=1$，$x>0$；$u(x)=0$，$x\leqslant 0$。显然，$t(\mathbf{y})$为观测到y_k为正的个数。

由于$1/2<Q_1^+<1$，所以$Q_1^+/(1-Q_1^+)>1$，即$L(y)$是关于$t(\mathbf{y})$的单调递增函数，由此，恒虚警概率为α检验由下式给出

$$\tilde{\delta}_s(\mathbf{y})=\begin{cases}1, & t(\mathbf{y})>\tau\\ \gamma, & t(\mathbf{y})=\tau\\ 0, & t(\mathbf{y})<\tau\end{cases} \tag{3.5.14}$$

这里γ和τ分别是虚警概率为α下的随机数和阈值。易知$t(\mathbf{Y})$是在H_1'下服从(n, Q_1^+)和在H_0'下服从$(n, 1/2)$的二项随机变量。给定虚警概率α，阈值是满足下式的最小整数

$$2^{-n}\sum_{k=\tau}^{n}\frac{n!}{(n-k)!k!}\leqslant\alpha \tag{3.5.15}$$

随机常数是

$$\gamma=\frac{\alpha-2^{-n}\sum_{k=\tau+1}^{n}\frac{n!}{(n-k)!k!}}{2^{-n}\frac{n!}{(n-\tau)!\tau!}} \tag{3.5.16}$$

对任意$Q_0\in\mathcal{P}_0$，$t(\mathbf{Y})$服从参量为$(n, 1/2)$的二项分布；对$\mathcal{P}_0$类中的所有分布，

式(3.5.14)的虚警概率都是 α。因此，假设检验(3.5.14)是非参数检验。这也说明式(3.5.14)是显著性水平为 α 的最大势检验。另外，式(3.5.14)的检验不取决于 Q_1 的选择，因此具有一致最大势。

因为 $t(\boldsymbol{Y})$ 是在 H_1 下服从 $(n,\ p)$ 的二项分布，式(3.5.14)的检测概率为

$$P_{\mathrm{D}}=\sum_{k=\tau+1}^{n}\frac{n!}{(n-k)!k!}p^{k}(1-p)^{n-k}+\gamma\frac{n!}{(n-\tau)!\tau!}p^{\tau}(1-p)^{n-\tau} \tag{3.5.17}$$

由式(3.5.17)可以看出，尽管在 H_0 下 $\tilde{\delta}_s$ 的虚警概率是常量，但是其检测概率仍取决于 $P\in\mathcal{P}_1$，因此也就受到 P 的选择约束。显然，随着 p 从 1/2 变到 1，P_{D} 从 α 单调增大到 1。

假设检验式(3.5.14)只使用观测值的符号特征，这就是众所周知的符号检测。虽然符号检测是显著性水平为 α 的一致最大势(UMP)检验。不过，如果已知两种假设下观测值的精确分布，利用似然比进行检测能获得更好的检测结果。换言之，给定与类 $\mathcal{P}_1$ 相对的 $P\in\mathcal{P}_0$ 后，符号检测并非显著性水平为 α 的一致最大势检验。因此，研究检测器在观测值分布信息未知的情况下相对于部分信息已知的性能损失就显得非常重要。

为了更好地解答这个问题，考虑利用 3.3 节中所介绍的 Pitman 渐近相对有效性(AER)来进行渐近 $n\to\infty$ 分析。根据定义，一个检测器相对另一个检测的渐近有效性是通过利用样本趋于无穷时达到相同性能所需要的相对样本数来衡量。

为了更好地利用渐近相对有效性来分析符号检测，需要建立一个特殊的检验模型来作为参考。为此，再次考虑如下模型

$$\begin{aligned}H_0: Y_k &= N_k, & k&=1,2,\cdots,n\\ H_1: Y_k &= N_k+\theta, & k&=1,2,\cdots,n\end{aligned} \tag{3.5.18}$$

其中 $N_1,\ N_2,\ \cdots,\ N_n$ 是均值为 0 的独立同分布序列。与上面讨论的结果一致，式(3.5.18)可归结为独立同分布加性噪声中的恒定信号检测问题。

如果假设式(3.5.18)中的噪声服从方差 σ^2 未知的正态分布 $\mathcal{N}(0,\ \sigma^2)$，则在无偏判决中，显著性水平为 α 检测的 UMP 可由下式给出

$$\tilde{\delta}_t(y)=\begin{cases}1, & \dfrac{\overline{y}}{[\overline{s^2}]^{1/2}}>\tau\\ \gamma, & \dfrac{\overline{y}}{[\overline{s^2}]^{1/2}}=\tau\\ 0, & \dfrac{\overline{y}}{[\overline{s^2}]^{1/2}}<\tau\end{cases} \tag{3.5.19}$$

其中 $\overline{y}$ 和 $\overline{s^2}$ 分别是样本均值和方差，定义为 $\overline{y}\triangleq\dfrac{1}{n}\sum_{k=1}^{n}y_k$ 和 $\overline{s^2}\triangleq\dfrac{1}{n}\sum_{k=1}^{n}(y_k-\overline{y})^2$ 。

式(3.5.19)所描述的假设检验也称为 t 检验。这个检验不但是式(3.5.18)在高斯条件下的 UMP，而且通过恰当选择门限 $\tau=\Phi^{-1}(1-\alpha)/\sqrt{n}$ 和随机值 γ 后，只要噪声具有零均值且方差有限，它就是虚警概率 $P_{\mathrm{F}}=\alpha$ 的渐近非参数检测器。因此，虚警概率可由下式给出

$$P_{\mathrm{F}}(\widetilde{\delta}_t)=P_0[\overline{Y}/(\overline{S^2})^{1/2}>\tau]+\gamma P_0[\overline{Y}/(\overline{S^2})^{1/2}=\tau]$$

$$=P_0\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{n}Y_k/(\overline{S^2})^{1/2}>\Phi^{-1}(1-\alpha)\right]+\gamma P_0\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{n}Y_k/(\overline{S^2})^{1/2}=\Phi^{-1}(1-\alpha)\right] \tag{3.5.20}$$

根据弱大数定理，$\overline{S^2}$依概率收敛于 $\mathrm{var}(N_1)$；又根据中心极限定理，H_0 下 $n^{-1/2}\sum_{k=1}^{n}Y_k/(\overline{S^2})^{1/2}$依分布收敛于服从$\mathcal{N}(0,1)$分布的随机变量。因此

$$\lim_{n\to\infty}P_{\mathrm{F}}(\widetilde{\delta}_t)=\frac{1}{\sqrt{2\pi}}\int_{\Phi^{-1}(1-\alpha)}^{\infty}\mathrm{e}^{-x^2/2}\mathrm{d}x=\alpha \tag{3.5.21}$$

式(3.5.21)表明 t 检验是渐近非参数检验。(注意，因为分布的极限连续，所以与 γ 无关)。

从上面的分析可知，t 检验是高斯噪声下式(3.5.18)的最优检验，也是有限方差噪声下的渐近非参数检验。既然第一个问题与检验式(3.5.3)的子集对应，在对噪声增加零中值约束后，第二个问题也与检验式(3.5.3)的子集对应。所以对于后一种情况，对比研究符号检测和 t 检验就非常有意义。

如果假设式(3.5.3)的噪声概率分布函数 f 的均值为 0，方差 $\sigma^2<\infty$，且分布函数在零处连续，直接利用 Pitman-Noether 定理，式(3.5.18)所描述的 t 检验相对于符号检测的渐近有效性为

$$\mathrm{ARE}_{s,t}=4\delta^2 f^2(0) \tag{3.5.22}$$

以高斯噪声为例，已知其概率密度函数 f 是服从$\mathcal{N}(0,\sigma^2)$的正态分布，则式(3.5.22)变为

$$\mathrm{ARE}_{s,t}=4\sigma^2\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^2=\frac{2}{\pi}\approx 0.64$$

即 t 检验需要的样本是同等符号检测所需样本数的 64%。对拉普拉斯噪声$[f(x)=(\alpha/2)\mathrm{e}^{-\alpha|x|}]$而言，代入 $\sigma^2=2/\alpha^2$ 后，有

$$\mathrm{ARE}_{s,t}=\frac{8}{\alpha^2}\left(\frac{\alpha}{2}\right)^2=2$$

即 t 检验所需样本数是符号检测的 2 倍。需要说明的是，从渐近有效性来看，符号检测在拉普拉斯噪声背景下是最优的。

事实上[详见 Kendall 和 Stuart(1961)]，对于任意具有对称单峰形式的密度函数[例如，对于$|x_2|>|x_1|$，有 $f(x)=f(-x)$和 $f(|x_1|)>f(|x_2|)$成立]，$\mathrm{ARE}_{s,t}$渐近有效性满足如下不等式

$$\mathrm{ARE}_{s,t}\geqslant 1/3 \tag{3.5.23}$$

因此，在这些条件下，t 检验需要的样本数至少为同等符号检测的 1/3。因为没有一致上界，当噪声分布很宽时，符号检测比 t 检验更有效。此外，就大量不同类型的分布而言，符号检测都是非参数化的，而 t 检验仅是某些分布下的渐近非参数化假设检验。符号检测的另一个突出优点是计算简单，在信号检测中经常用来代替最优方法。尤其是在雷达恒虚警检测中，符号检测和 t 检验都得到了广泛应用(此时符号检测也称为二进制积累器/累加器)。

2. 秩检验

从上面的讨论可以看出，相比其他需要观测数据先验信息的其他检验方法而言，即使

在最坏情况下，符号检验也是相当有效的。不过，如果能够利用观测数据更多的先验信息，就能在保留符号检验非参数特性的同时大幅度提高最坏条件下的检测性能。

例如，将式(3.5.13)所描述的符号检测中的统计量 $t(\mathbf{y}) = \sum_{k=1}^{n} u(y_k)$ 用带权重的 $\sum_{k=1}^{n} \lambda_k u(y_k)$ 替代，其中 λ_k 为样本 y_k 按照绝对值增大的顺序重新排列后的序数。即假设将 y_1，…，y_n 按照 y_{k_1}，…，y_{k_n} 来排列(其中 $|y_{k_1}| \leqslant |y_{k_2}| \leqslant \cdots \leqslant |y_{k_n}|$)后，对统计量进行门限检测

$$t_w(\mathbf{y}) = \sum_{i=1}^{n} iu(y_{k_i}) \tag{3.5.24}$$

这就是著名的 Wilcoxon 检验，它是秩检验的特例，依赖于每个样本在整个观测序列中的秩。

式(3.5.24)可改写为

$$t_W(\mathbf{y}) = \sum_{k=1}^{n} \sum_{j=1}^{k} u(y_k + y_j) \tag{3.5.25}$$

式(3.5.25)的具体推导过程留作练习。显然，在观测数据 Y_1，…，Y_n 满足独立同分布条件和边缘分布对称[对于所有实数 b 都有 $F_{Y_k}(b) = 1 - F_{Y_k}(-b)$]时，Wilcoxon 秩检验是非参数化的假设检验。相比中值为 0 的一类分布(符号检测仍是非参数化的)而言，此时的概率分布范围要小得多。针对式(3.5.18)所描述的假设检验问题，根据 Pitman-Noether 定理，Wilcoxon 秩检验相对 t 检验的渐近有效性为

$$\mathrm{ARE}_{w,t} = 12\sigma^2 \left| \int_{-\infty}^{\infty} f^2(x)\mathrm{d}x \right|^2 \tag{3.5.26}$$

在式(3.5.26)中需要假设噪声具有对称分布特性。在高斯噪声[$N_k \sim \mathcal{N}(0, \sigma^2)$]下，有 $\mathrm{ARE}_{w,t} = 3/\pi = 0.955$，即 Wilcoxon 检验在高斯背景下近似最优。在拉普拉斯噪声条件下，$\mathrm{ARE}_{w,t} = 1.5$，即相比符号检验而言存在 25%的渐近有效性损失。更一般地，在噪声对称分布和 $\int_{-\infty}^{\infty} f^2(x)\mathrm{d}x$ 约束下，最小化 $\int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x = \sigma^2$ 后[详见 Kendall 和 Stuart(1961)]，有

$$\mathrm{ARE}_{w,t} \geqslant 0.864 \tag{3.5.27}$$

显然，Wilcoxon 检验相对 t 检验而言，渐近有效性不低于 86.4%。另外，因为无法限制方差，所以 $\mathrm{ARE}_{w,t}$ 没有确定的上界，即 Wilcoxon 检验要比 t 检验有效得多。不过，Wilcoxon 检验需要存储所有样本后才能计算统计量，不利于工程实现。

相比 t 检验而言，可以通过采用比 Wilcoxon 检验更复杂的秩检验来获得更好的性能。例如 Fisher-Yates 检验，也称归一化评分检验，它对应的统计量为

$$t_{\mathrm{FY}}(\mathbf{y}) = \sum_{i=1}^{n} h_n[iu(y_{k_i})] \tag{3.5.28}$$

和 Wilcoxon 检验一样，y_{k_1}，…，y_{k_n} 是有序样本，函数 h_n 定义式为

$$h_n(i) = \begin{cases} 0, & i = 0 \\ E\{X_{(i)}\}, & i = 1,\cdots,n \end{cases} \tag{3.5.29}$$

其中 $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ 是服从标准正态分布的随机变量 X_1，X_2，…，X_n 对应的有序

值。相对于 t 检验，Fisher-Yates 检验的效率满足

$$\mathrm{ARE}_{\mathrm{FY},t} \geqslant 1$$

可见，Fisher-Yates 检验的效率总不低于 t 检验。不过，Fisher-Yates 检验的计算复杂度明显高于 t 检验。

有关秩检验更深入的讨论，建议读者参考 Hajek、Sidak(1967)和 Kendall(1948)的著作。

3. 双通道检验

在很多应用中，观测数据由两个或多个传感器(通道)同时采样获得。例如在声呐、地震学、射电天文等研究领域，常采用阵列传感器来采集信号。下面结合双通道检测实例介绍多通道检测的非参数化方法。

考虑包含 n 对独立随机变量的观测序列，如 $\boldsymbol{Y}=[(U_1, V_1), (U_2, V_2), \cdots, (U_n, V_n)]$，其中$(U_k, V_k)$，$k=1, 2, \cdots, n$ 互相独立。利用两个传感器检测是否存在随机信号是上述模型的特例，可建模为

$$\begin{aligned} H_0: & \begin{matrix} U_k = N_k \\ V_k = W_k \end{matrix}, \qquad k = 1,2,\cdots,n \\ H_1: & \begin{matrix} U_k = N_k + S_k \\ V_k = W_k + S_k \end{matrix}, \quad k = 1,2,\cdots,n \end{aligned} \tag{3.5.30}$$

其中$\{N_k\}_{k=1}^{n}$、$\{W_k\}_{k=1}^{n}$和$\{S_k\}_{k=1}^{n}$是服从独立同分布且相互独立的随机序列，边缘分布函数分别为 F_N、F_W 和 F_S。

因为随机变量相互独立，则(U_k, V_k)在 H_0 下的联合分布函数为

$$Q_0(u,v) = F_N(u)F_W(v) \tag{3.5.31}$$

在 H_1 下的联合分布函数为

$$Q_1(u,v) = \int_{-\infty}^{\infty} F_N(u-s)F_W(v-s)\,\mathrm{d}F_S(s) \tag{3.5.32}$$

如果联合分布函数 F_N、F_W 和 F_S 已知，结合式(3.5.31)和式(3.5.32)可计算出式(3.5.30)中最优检测所需的似然比。

例如，假设 F_N、F_W 和 F_S 都是双变量高斯分布，每个变量的均值都为零、方差分别为 $\mathrm{var}(N_k)=\mathrm{var}(W_k)=\sigma^2$ 和 $\mathrm{var}(S_k)=\sigma_s^2$，且两个变量的相关系数为 0。则 Q_0 是双变量高斯密度函数，两个分量的均值均为 0，方差均为 σ^2，相关系数为 0；Q_1 也是双变量高斯密度，两个分量的均值均为 0，但方差都为 $\sigma^2/(1-\rho)$，相关系数为 ρ，其中 $\rho \triangleq \sigma_s^2/(\sigma^2+\sigma_s^2)$。对于已知的 σ^2，虚警概率为 α 的 UMP 检验由能量检测器给出，即

$$\delta_{\mathrm{ED}}(\boldsymbol{y}) = \begin{cases} 1, & \sum\limits_{k=1}^{n}(u_k+v_k)^2 > \tau \\ \gamma, & \sum\limits_{k=1}^{n}(u_k+v_k)^2 = \tau \\ 0, & \sum\limits_{k=1}^{n}(u_k+v_k)^2 < \tau \end{cases} \tag{3.5.33}$$

门限 τ 的大小由虚警概率 α 决定。式(3.5.33)的性能分析方法与3.2节单通道辐射计的分析方法类似。

在实际应用中，联合分布函数 F_N、F_W 和 F_S 通常未知，有必要寻求非最优检测器。例如常用的极性重合相关器(PCC)，它的判决形式为

$$\delta_{\mathrm{PCC}}(\boldsymbol{y})=\begin{cases}1, & \sum_{k=1}^{n}u(u_k v_k)>\tau\\ \gamma, & \sum_{k=1}^{n}u(u_k v_k)=\tau\\ 0, & \sum_{k=1}^{n}u(u_k v_k)<\tau\end{cases} \tag{3.5.34}$$

其中 u 代表单位阶跃函数，即

$$u(x)=\begin{cases}1, & x>0\\ 0, & x\leqslant 0\end{cases} \tag{3.5.35}$$

检测器根据观察数据"极性重合"的次数进行判决，也就是根据两个通道的数据符号相同的次数进行判决。根据式(3.5.30)的模型，H_1 假设下两个通道的信号相同，H_0 假设下两个通道均无信号。因此，H_1 下极性重合的次数应该更大。显然，极性重合的次数可作为判断信号是否存在的合理途径。

无论是 H_0 假设还是 H_1 假设，PCC 的统计量 $t_{\mathrm{PCC}}(Y)\triangleq\sum_{k=1}^{n}u(U_k,V_k)$ 都是 n 个独立同分布伯努利随机变量的和。因此，在 H_j 假设下，$t_{\mathrm{PCC}}(Y)$ 是服从(n，λ_j)的二项式随机变量，其中 $\lambda_j=P_j(U_1V_1\geqslant 0)$，$j=0$，1。可以写为

$$\lambda_j=P_j(U_1\geqslant 0,V_1\geqslant 0)+P_j(U_1\leqslant 0,V_1\leqslant 0)$$

简化后有

$$\lambda_j=1-Q_j(0,\infty)-Q_j(\infty,0)+2Q_j(0,0) \tag{3.5.36}$$

结合式(3.5.31)和式(3.5.32)，有

$$\lambda_0=1-F_N(0)-F_W(0)+2F_N(0)=\frac{1}{2}+2\left[F_N(0)-\frac{1}{2}\right]\left[F_W(0)-\frac{1}{2}\right] \tag{3.5.37}$$

和

$$\begin{aligned}\lambda_1&=1-\int_{-\infty}^{\infty}F_N(-s)\mathrm{d}F_S(s)-\int_{-\infty}^{\infty}F_W(-s)\mathrm{d}F_S(s)\\&\quad+2\int_{-\infty}^{\infty}F_N(-s)F_W(-s)F_S(s)\\&=\frac{1}{2}+2\int_{-\infty}^{\infty}\left[F_N(s)-\frac{1}{2}\right]\left[F_W(s)-\frac{1}{2}\right]\mathrm{d}F_S(s)\end{aligned} \tag{3.5.38}$$

需要注意的是，只要任意一个噪声过程的中值为0，即 $F_N(0)=1/2$ 或者 $F_W(0)=1/2$，就有 $\lambda_0=1/2$ 成立。进而，在至少有一个中值为0的一类噪声分布下，PCC 都是非参数化的检测器。这是因为在这类噪声分布下，t_{PCC} 服从(n，1/2)分布，且通过合理设置 τ 和 γ 可实现设定的虚警概率 α。

对比 PCC 和高斯信道对应的最优检测器是非常有意义的。假设 F_N 和 F_W 的概率密度函数分别为 f_N 和 f_W，同时假设 f_N 和 f_W 在 0 处连续且二阶矩和四阶矩为有限值，针对式(3.5.30)所描述的假设检验模型，δ_{PCC}相对于 δ_{ED}的渐近有效性可表示为

$$\mathrm{ARE}_{\mathrm{PCC,ED}} = [\gamma_N^4 + \gamma_W^4 + 4\sigma_N^2\sigma_W^2 - \sigma_N^2 - \sigma_W^4] f_N^2(0) f_W^2(0)$$

其中

$$\sigma_N^2 \triangleq E\{N_1^2\}, \quad \sigma_W^2 \triangleq E\{W_1^2\}, \quad \gamma_N^2 \triangleq E\{N_1^4\}, \quad \gamma_W^4 \triangleq E\{W_1^4\}$$

以同分布的高斯信道为例，即

$$N_1 \sim \mathcal{N}(0,\sigma^2), \quad W_1 \sim \mathcal{N}(0,\sigma^2)$$

可得

$$\sigma_N^2 = \sigma_W^2 = \sigma^2$$

进而可计算出

$$\mathrm{ARE}_{\mathrm{PCC,ED}} = \frac{2}{\pi^2} = 0.202$$

类似地，对于相同分布的拉普拉斯信道而言，即 $f_{N_1}(x) = f_{W_1}(x) = \frac{\alpha}{2}\mathrm{e}^{-\alpha|x|}$，代入 $\sigma_N^2 = \sigma_W^2 = 2/\alpha^2$ 和 $\gamma_N^4 = \gamma_W^4 = 24/\alpha^2$ 后，有

$$\mathrm{ARE}_{\mathrm{PCC,ED}} = 3.5$$

在高斯信道下，PCC 相比功率检测器而言要差得多；但在拉普拉斯信道下，PCC 却能获得更好的性能。在高斯信道下，PCC 以降低效率为代价，实现处理的简单化和非参数化检测。与单信道情况类似，采用基于秩的更复杂的检测结构后，双通道或多通道检测也能在保留非参数性能的同时大幅度提高性能。

式(3.5.30)给出了双信道检测模型的一般情况。双通道检验的另一类问题是在参考噪声源中检测信号，即假设检验问题可描述成

$$\begin{aligned} &H_0: \begin{matrix} U_k = N_k \\ V_k = W_k \end{matrix}, \qquad k = 1,2,\cdots,n \\ &H_1: \begin{matrix} U_k = N_k \\ V_k = W_k + S_k \end{matrix}, \quad k = 1,2,\cdots,n \end{aligned} \tag{3.5.39}$$

其中$\{N_k\}_{k=1}^n$、$\{W_k\}_{k=1}^n$是相互独立且具有相同边缘分布的独立同分布噪声序列，$\{S_k\}_{k=1}^n$是与噪声独立的信号序列。增加 U 通道的一个噪声样本后，该问题类似于常见的 V 通道单样本检测问题，在不存在检测信号的时段，两组噪声样本可由一个通道采集，因此可存储下来作为后续处理的附加样本。为了简便起见，在 3.5.39 中假设每个通道下的样本数一致，尽管这一假设在实际处理中并非必要。

针对式(3.5.39)所描述的假设检验问题，有多种判决准则可用来检测信号。在独立假设下，如果噪声的统计特性精确已知，则仅含噪声的采样通道无法用于检测，原因在于似然比中与 U_k有关的所有量都消失了。因此，只有在噪声分布未知时，不含信号的信道才是有用的。

只有当信号是正常数时，双样本形式的 Wilcoxon 检验才是式(3.5.39)的一种有效检验方法，它基于统计量

$$\sum_{k=1}^{n} r_k \tag{3.5.40}$$

其中，r_k 是将所有观测值 u_1，…，u_n 和 v_1，…，v_n 按照递增顺序合并排列后 v_k 的秩。这种检验方法也称为 Mann-Whitney 检验。与式(3.5.40)等价的统计量是

$$\sum_{k=1}^{n}\sum_{j=1}^{n} u(v_k - u_j)$$

由上式可知，噪声对称分布下的 Mann-Whitney 检测器是非参数化检验。与单通道的 Wilcoxon 检验类似，与式(3.5.40)的检测器相比，采用双信道检验是非常有效的。

Mann-Whitney 检验可归结到 Kolmogorov-Smirnov 检验的范畴，此类检验基于下列函数表示的泛函

$$\hat{F}_V(x) - \hat{F}_U(x), \quad -\infty < x < \infty$$

其中$\hat{F}_V(x)$和$\hat{F}_U(x)$分别是样本$\{u_k\}_{k=1}^{n}$和$\{v_k\}_{k=1}^{n}$的经验分布函数，即$\hat{F}_V(x)$定义为

$$\hat{F}_V(x) = \frac{1}{n}\sum_{k=1}^{n} u(x - v_k) \quad -\infty < x < \infty \tag{3.5.41}$$

$\hat{F}_U$ 具有相同的定义[⊖]。随着 $n \to \infty$，$\hat{F}_V(x)$和$\hat{F}_U(x)$分别收敛于 U_k 和 V_k 的边缘分布，其差函数$\hat{F}_V - \hat{F}_U$ 可用来判断$\{U_k\}_{k=1}^{n}$和$\{V_k\}_{k=1}^{n}$是否服从相同的分布。除 Mann-Whitney 检验之外，将统计量 $\sup\limits_{-\infty<x<\infty}[\hat{F}_V(x) - \hat{F}_U(x)]$或 $\sup\limits_{-\infty<x<\infty}|\hat{F}_V(x) - \hat{F}_U(x)|$与门限进行比较的检测器也属于非参数检验范畴。在$\{U_k\}_{k=1}^{n}$和$\{V_k\}_{k=1}^{n}$服从相同边缘分布时，它们是非参数化检验，且在一定虚警概率范围内的检测性能优于 Mann-Whitney 检验。在非参数检测的相关理论和应用方面更详尽的描述，可参考 Kassam 和 Thomas(1980)的著作。

3.5.2 稳健检测

在 3.2 节中研究了观测数据统计特性精确已知下的检测系统设计问题；在 3.5.1 中研究了几乎不依赖观测值统计特征的检测系统。除去这两种极端情况，还存在精确的统计模型在名义上可用，但实际上有偏差的情况。

例如，针对独立同分布序列 Y_1，…，Y_n，考虑在两个可能边缘分布 P_0 和 P_1 之间的假设检验问题。假设密度为 p_0 和 p_1，则基于似然比的最优检验为

$$L(\mathbf{y}) = \prod_{k=1}^{n} \frac{p_1(y_k)}{p_0(y_k)} \tag{3.5.42}$$

注意，不论是 $p_1(y_k) \gg p_0(y_k)$，还是 $p_0(y_k) \gg p_1(y_k)$，似然比对观测序列值都非常敏感。由于 $p_1(y_k) \gg p_0(y_k)$在 H_1 下出现的可能性远大于它在 H_0 下出现的可能性(反之亦然)，因此，似然比的敏感性通常也就默认为判决器的特性之一。

进一步假设实际的边缘分布不精确等于 P_0 和 P_1，其近似表达形式为 P_0 或 P_1。例如，假设实际分布有如下形式

$$(1-\varepsilon)P_j + \varepsilon M_j, \quad j = 0,1 \tag{3.5.43}$$

这里 P_0 和 P_1 是名义上的分布，M_0 和 M_1 未知，且为任意“污染”分布，ε 是 0～1 之间的

⊖ 对每个 x，$\hat{F}_V(x)$是 v_k 大于等于 x 的个数除以 v_k 的总数。

数，表示此模型中的不确定度。例如，在通信信道或雷达传播环境中，外部干扰只是在小部分时间内出现，或者脉冲噪声(闪电等)以概率 ε 出现。同样地，传感器间歇性失效、测量或数据记录错误也可按这种方法建模。式(3.5.43)的核心是用 M_j 来表征估计模型中完全未知的部分。

假设 $p_1(y_k)/p_0(y_k)$ 是关于 y_k 的无界函数。由于 M_0 任意选取，则其可取遍所有 $p_1(y_k)\gg p_0(y_k)$ 的区域。如果仍然采用标称似然比进行判决，那么可能会增大虚警概率。正如下文所述，模型误差会使得虚警概率(和整体错误概率)大致在 $(1-(1-\varepsilon)^n)$ 量级，且随着 n 的增大而增大，在 $n=1$ 时等于 ε。很多检测系统的虚警概率都设计在 $10^{-4}\sim10^{-6}$ 范围，因此，即使存在 1% 的模型不确定性，都可能会影响系统的虚警概率。H_1 下也存在类似的问题，即如果 $p_1(y_k)/p_0(y_k)$ 可无限趋近于 0，在 $p_1(y_k)\ll p_0(y_k)$ 的区域内遍取 M_1 后会导致似然比低于判决门限。[⊖]

尽管上面的讨论是定性的，但也正是因为 $p_1(y_k)/p_0(y_k)$ 可无穷大，也可无限趋近 0，导致了似然比检验缺乏稳健性。特别是模型中相对小的偏差都可能会导致明显的性能损失。能否采用其他途径来增强稳健性呢？显然，根据前面的分析，如果将 p_1/p_0 用上下都有界的形式替换后，就能增强似然比检验的稳健性。换言之，假设将乘积 $\prod\limits_{k=1}^{n}p_1(y_k)/p_0(y_k)$ 中的 $l\triangleq p_1/p_0$ 替换为

$$[l]_a^b(y)=\begin{cases}b, & l(y)>b\\ l(y), & a\leqslant l(y)\leqslant b\\ a, & l(y)<a\end{cases}\tag{3.5.44}$$

其中 $0<a<b<\infty$。通过选取恰当的 a 和 b，基于 $\prod\limits_{k=1}^{n}[l]_a^b(y_k)$ 的检验不会面临 l 无界的问题。然而，采用式(3.5.44)的形式后，名义上的最优性会随之散失，这是因为当 $l(y)$ 非常大或接近 0 时，观测所携带的信息量最大。

在不确定模型下寻求性能最优的检验准则是非常有意义的。事实上，上述关于似然比检验稳健性的讨论为设计稳健的判决准则提供了依据。这些讨论是以不确定模型的最差性能为出发点，因此，在标称模型的某个恰当的邻域内用最差性能代替错误概率就是非常合理的设计准则[见式(3.5.43)]，进而通过优化给定准则下的系统性能来设计判决准则。

回顾一下，如果序列 P_0 是 Y_k 的真实边缘分布，检验 δ 的虚警概率为

$$P_{\mathrm{F}}(\delta,P_0)=\int_{\Gamma}\delta(\boldsymbol{y})\Big[\prod_{k=1}^{n}p_0(y_k)\Big]\mu(\mathrm{d}\boldsymbol{y})\tag{3.5.45}$$

其中 p_0 是所对应的边缘概率密度函数。类似地，当 P_1 是准确的边缘分布时，漏警概率为

$$P_{\mathrm{M}}(\delta,P_1)=\int_{\Gamma}[1-\delta(\boldsymbol{y})]\Big[\prod_{k=1}^{n}p_1(y_k)\Big]\mu(\mathrm{d}\boldsymbol{y})\tag{3.5.46}$$

为简单起见，假设代价都相同，则常用的二元假设检验判据是：

1) $\min\limits_{\delta}[\pi_0P_{\mathrm{F}}(\delta,\ P_0)+\pi_1P_{\mathrm{M}}(\delta,\ P_1)]$(贝叶斯)

⊖ 即降低了检测概率。——译者注

2) $\min\limits_{\delta}[\max\{P_F(\delta,\ P_0),\ P_M(\delta,\ P_1)\}]$(极小化极大值)

3) $\min\limits_{\delta} P_M(\delta,\ P_1)$, s. t. $P_F(\delta,\ P_0)\leqslant\alpha$(尼曼-皮尔逊)

假设 Y_k 的真实边缘分布与名义边缘分布 P_0 或 P_1 存在差异，但在 P_0 和 P_1 的邻域 $\mathcal{P}_0$、$\mathcal{P}_1$ 内。在 1)～3)中，将 $P_F(\delta,\ P_0)$和 $P_M(\delta,\ P_1)$用最差性能值代替，即

$$P_F(\delta,\mathcal{P}_0)\triangleq\sup_{P\in\mathcal{P}_0}P_F(\delta,P) \tag{3.5.47}$$

和

$$P_M(\delta,\mathcal{P}_1)\triangleq\sup_{P\in\mathcal{P}_1}P_M(\delta,P) \tag{3.5.48}$$

检验问题的等价形式为：

1′) $\min\limits_{\delta}[\pi_0 P_F(\delta,\ \mathcal{P}_0)+\pi_1 P_M(\delta,\ \mathcal{P}_1)]$

2′) $\min\limits_{\delta}[\max\{P_F(\delta,\ \mathcal{P}_0),\ P_M(\delta,\ \mathcal{P}_1)\}]$

3′) $\min\limits_{\delta}P_M(\delta,\ \mathcal{P}_1)$, s. t. $P_F(\delta,\ \mathcal{P}_0)\leqslant\alpha$

在所有可能的检验中，1′)～3′)的最差解在邻域$\mathcal{P}_0$ 和$\mathcal{P}_1$ 内的代价最小。当然，在邻域设置过大时，最大化最差性能的判决可能会过于保守。然而，此处认为$\mathcal{P}_0$ 和$\mathcal{P}_1$ 是名义模型的小邻域，且目的在于增强该邻域内的稳健性。事实证明，对于式(3.5.43)的不确定模型，1′)～3′)的解能够实现预期的稳健检测目的。

虽然 Huber 和 Strassen(1973)，提供了 1′)～3′)的一般性求解方法，但其稍显复杂，在此不予讨论。作为补充，本节将集中讨论式(3.5.43)(也称为 ε 污染混合模型)在特殊的不确定邻域下的求解方法。这种解法由 Hubert 在 1965 年提出，并首次将 1′)～3′)作为稳健检测的设计准则。

事实证明，1′)～3′)的解是在 $Q_0\in\mathcal{P}_0$ 和 $Q_1\in\mathcal{P}_1$ 内用最不利分布代替 P_0 和 P_1 后 1)～3)的最优解，其中 Q_0 和 Q_1 的密度为

$$q_0(y_k)=\begin{cases}(1-\varepsilon)p_0(y_k), & p_1(y_k)<c''p_0(y_k)\\ \dfrac{1-\varepsilon}{c''}p_1(y_k), & p_1(y_k)\geqslant c''p_0(y_k)\end{cases} \tag{3.5.49a}$$

和

$$q_1(y_k)=\begin{cases}(1-\varepsilon)p_1(y_k), & p_1(y_k)>c'p_0(y_k)\\ c'(1-\varepsilon)p_0(y_k), & p_1(y_1)\leqslant c'p_0(y_k)\end{cases} \tag{3.5.49b}$$

其中，$0<c'<1<c''<\infty$是使得 Q_1 和 Q_2 的概率为 1 的常数，即

$$(1-\varepsilon)[P_0(l(Y_k)<c'')+P_1(l(Y_k)\geqslant c'')/c'']=1 \tag{3.5.50a}$$

和

$$(1-\varepsilon)[P_1(l(Y_k)>c')+c'P_0(l(Y_k)\leqslant c')]=1 \tag{3.5.50b}$$

似然比是 1)～3)的解，因此 1′)～3′)的解也是 Q_1 和 Q_2 的似然比；换句话说，它们基于似然比 $\prod\limits_{k=1}^{n}q_1(y_k)/q_0(y_k)$。结合式(3.5.49)，有

$$\frac{q_1(y_k)}{q_0(y_k)}=\begin{cases}c', & l(y_k)<c'\\ l(y_k), & c'\leqslant l(y_k)\leqslant c''\\ c'', & l(y_k)>c''\end{cases} \tag{3.5.51}$$

从而式 1′)～3′)的解是基于下式的门限检测

$$\prod_{k=1}^{n}[l]_{c'}^{c''}(y_k) \tag{3.5.52}$$

至此，得到了稳健检验的解析形式。就优化而言，上述判决形式在求解 1′)～3′)时具有独特作用。特别是，对于基于似然比[式(3.5.52)]的任意门限检测 δ_R 而言，有下式成立

$$P_F(\delta_R,\mathcal{P}_0)=P_F(\delta_R,Q_0) \tag{3.5.53a}$$

和

$$P_M(\delta_R,\mathcal{P}_1)=P_M(\delta_R,Q_1) \tag{3.5.53b}$$

这表明 δ_R 的最坏性能实际上是在分布对(Q_0，Q_1)上的最优性能。这使得可以简单地通过估计 δ_R 在边缘分布 Q_0 和 Q_1 上的错误概率来计算$\mathcal{P}_0$ 和$\mathcal{P}_1$ 类上 δ_R 的错误概率的上界。

备注

1) 从某种意义上说，最不利分布 Q_0 和 Q_1 是在 $Q_0\in\mathcal{P}_0$ 和 $Q_1\in\mathcal{P}_1$ 约束下形状尽可能相近的分布。能够证明，如果 $P_0\neq P_1$ 且 ε 足够小(为正)，则方程(3.5.50a)的解满足 $0<c'<1<c''<\infty$。这意味着 $Q_0\neq Q_1$ 和 δ_R 是有意义的检验判决。另一方面，如果 ε 太大，则$\mathcal{P}_0$ 和$\mathcal{P}_1$ 将会重叠，$c'=c''=1$ 且 $Q_0=Q_1$。在这种情况下，$q_1(y_k)/q_0(y_k)\equiv 1$ 和式 1′)～3′)的解忽略了观测值。因此，如果$\mathcal{P}_0$ 和$\mathcal{P}_1$ 的邻域太大以至于发生重叠，那么 1′)～3′)就不再是可行的设计准则。不过，如上所述，该设计准则仅针对 ε 较小的情况。

2) 即使 ε 非常小，分布采用 l 和$[l]_{c'}^{c''}$进行判决后，差异也十分明显。以 $\pi_0=\pi_1=1/2$ 时的贝叶斯公式 1)为例，门限具有一致性且可进行任意的随机化处理。假设 $0<l(y_k)<\infty$ 和 $\sup_{y_k\in R}l(y_k)=\infty$，因为 M_0 可取遍所有的可能性，那么对于依赖似然比 l 的判决 δ_0 而言，任意观测均可导致概率为 ε 的虚警。在 n 个独立观测下，有

$$P_F(\delta_0,\mathcal{P}_0)\geqslant 1-(1-\varepsilon)^n \tag{3.5.54}$$

类似地，如果 $\inf_{y_k\in\mathbf{R}}l(y_k)=0$，则在$\mathcal{P}_1$ 下 δ_0 的漏警概率为

$$P_M(\delta_0,\mathcal{P}_1)\geqslant 1-(1-\varepsilon)^n \tag{3.5.55}$$

综上，在$\mathcal{P}_0$ 和$\mathcal{P}_1$ 上，最坏情况下 δ_0 的平均错误概率为

$$\sup P_e(\delta_0)=\frac{1}{2}P_F(\delta_0,\mathcal{P}_0)+\frac{1}{2}P_M(\delta_0,\mathcal{P}_1)\geqslant 1-(1-\varepsilon)^n$$

对于任意 $\varepsilon>0$，有 $\lim\limits_{n\to\infty}(1-\varepsilon)^n=0$ 成立，即

$$\lim_{n\to\infty}[\sup P_e(\delta_0)]=1 \tag{3.5.56}$$

这意味着 δ_0 的性能可以任意差。实际上，它可能比随机猜测的情况更差，这是因为猜测引起的错误概率是 1/2。

与此相反，考虑基于 q_1/q_0 的检验 δ_R。根据性质式(3.5.50a)，有

$$\sup P_e(\delta_R)=\frac{1}{2}P_F(\delta_R,Q_0)+\frac{1}{2}P_M(\delta_R,Q_1) \tag{3.5.57}$$

代入切诺夫界[式(3.3.18)]，有

$$\sup P_e(\delta_R)\leqslant\frac{1}{2}\left[\int[q_0q_1]^{1/2}\right]^n \tag{3.5.58}$$

正如 3.3 节中所提到的，如果 $Q_0\neq Q_1$，有 $\int[q_0q_1]^{1/2}<1$ 成立，则

$$\lim_{n\to\infty}[\sup P_e(\delta_R)]=0 \tag{3.5.59}$$

综合式(3.5.56)和式(3.5.59)易知：按照最差性能准则，$P_e(\delta_0)$以指数形式收敛于1，而$P_e(\delta_R)$依指数收敛于0。尽管使$P_e(\delta_0)$收敛于1的情况较为极端，但却能说明非稳健似然比检验的局限。需要注意的是，在通常情况下，根据切诺夫界，$P_e(\delta_0)$也收敛于0。因此，在最坏情况下，δ_R达到了与通常情况相似的性能。

3）除了ε污染混合模型之外，由其他非确定性模型也能得到式1′)～3′)的解。事实上，允许不同假设下的混合参数ε存在变化后，ε污染混合模型也可推广到更一般的形式。例如，不同假设下的邻域可表示为$\mathcal{P}_0=\{P|\rho(P,P_0)\leqslant\varepsilon_0\}$和$\mathcal{P}_1=\{P|\rho(P,P_1)\leqslant\varepsilon_1\}$，其中$\rho$为概率分布的某种距离度量方式。在大多数情况下，式1′～3′的其他形式的解与ε污染混合模型在最不利分布下的解近似，且它们常常需要对似然比进行某种幅度限制。ε污染混合模型也可推广为时变模型，此时仅需将$\prod_{k=1}^{n}[l_k]_{c'}^{c''}(y_k)$替换为$\prod_{k=1}^{n}[l_k]_{c'_k}^{c''_k}(y_k)$，其中$l_k$是第$k$个样本名义上的似然比，而$c'_k$和$c''_k$是式(3.5.50a)在第$k$个名义分布上的解。

下面通过实例来说明以上结论的应用。

例3.5.1　限幅-相关器

重新考虑相干信号检测问题如下：

$$\begin{aligned}H_0:Y_k&=N_k, & k&=1,2,\cdots,n\\ H_1:Y_k&=N_k+\theta s_k, & k&=1,2,\cdots,n\end{aligned} \tag{3.5.60}$$

其中，s_1，…，s_n是已知的信号序列，N_1，…，N_n是独立同分布且服从$\mathcal{N}(0,1)$的噪声序列，θ为已知振幅(正实数)。显然，第k个样本对应的似然比为$l_k(y_k)=\exp[\theta s_k(y_k-\theta s_k/2)]$[见式(3.2.9)]，其对数似然比为

$$\log L(\mathbf{y})=\theta\sum_{k=1}^{n}s_k(y_k-\theta s_k/2) \tag{3.5.61}$$

由此，可得到图3.5.1所示的相关检测器。◀

实际情况下，式(3.5.60)所描述模型只是近似正确。因为噪声严格服从$\mathcal{N}(0,1)$分布的可能性很小，测量存在非线性，且接收信号为信号和噪声之和的假设也不完全准确。因此，为更切合实际，可将式(3.5.60)改为

$$\begin{aligned}H_0:Y_k&\sim(1-\varepsilon)P_0+\varepsilon M_0, & k&=1,2,\cdots,n\\ H_1:Y_k&\sim(1-\varepsilon)P_1^{(k)}+\varepsilon M_1^{(k)}, & k&=1,2,\cdots,n\end{aligned} \tag{3.5.62}$$

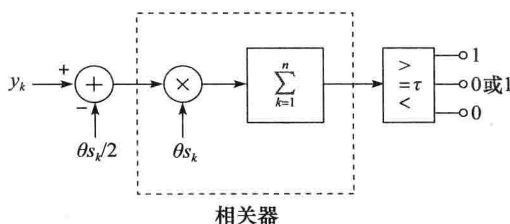


图3.5.1　高斯噪声下相关信号的最优检测器

其中P_0是$\mathcal{N}(0,1)$分布，$P_1^{(k)}$是$\mathcal{N}(\theta s_k,1)$分布，$M_0$和$M_1^{(k)}(k=1,2,\cdots,n)$为任意值($M_0$也可随$k$变化，但不会改变下面的结论)，$\varepsilon$的值在0～1之间。

考虑$l_k(y_k)=\exp[\theta s_k(y_k-\theta s_k/2)]$后，有$0<lk(y_k)<\infty$成立，即$\sup_{y_k\in\mathbf{R}}lk(y_k)=\infty$和$\inf_{y_k\in\mathbf{R}}lk(y_k)=0$。因此，与式(3.5.62)讨论的模型类似，相关检测器同样存在性能损失。在此条件下，Huber稳健似然比检验δ_R能获得较好性能。式(3.5.62)所描述的检验问题的稳健似然比$L_R(\mathbf{y})=\prod_{k=1}^{n}[l_k]_{c'_k}^{c''_k}(y_k)$的对数形式为

$$\log L_{\mathrm{R}}(\boldsymbol{y}) = \sum_{k=1}^{n} [\theta s_k(y_k - \theta s_k/2)]_{d_k'}^{d_k''} \tag{3.5.63}$$

其中，$d_k' \triangleq \log c_k'$，$d_k'' \triangleq \log c_k''$。与之对应的稳健检测器如图 3.5.2 所示。

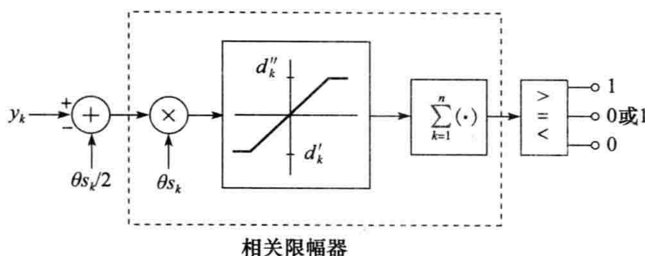


图 3.5.2 高斯噪声背景下的相关信号的稳健检测器

利用$\mathcal{N}(0, 1)$分布和似然比的对称性，可以得到 $d_k' = -d_k''$，且 d_k'' 就是下式的解

$$\Phi\left(\frac{d_k''}{\theta|s_k|} + \frac{\theta|s_k|}{2}\right) + \mathrm{e}^{-d_k''}\left[1 + \Phi\left(\frac{d_k''}{\theta|s_k|} - \frac{\theta|s_k|}{2}\right)\right] = (1-\varepsilon)^{-1} \tag{3.5.64}$$

其中 Φ 是$\mathcal{N}(0, 1)$的累积分布函数。显然，随着 ε 由 0 开始增大，d_k'' 从∞单调递减至 0。

对比图 3.5.1 与图 3.5.2 可知，图 3.5.2 在乘法器和累加器之间增加了将幅度限制在 $\pm d_k''$ 范围内的限幅器。该检测器也称为限幅-相关器，由 Martin 和 Schwartz 在 1971 年首次提出。只要 ε 小到可以避免 H_0 和 H_1 假设下的不确定集发生重叠，该检测器就是稳健的。对于固定的 ε，这相当于增大信号强度 θ 后能避免重叠。特别地，$|s_k|>0(k=1, 2, \cdots, n)$时，$H_0$ 和 H_1 假设下的不确定集不重叠要求 $\theta>\theta_\varepsilon/\min|s_k|$，其中 θ_ε 是式$(1-\varepsilon)\Phi(\theta_\varepsilon/2)=1/2$ 的解。例如，如果 $\varepsilon=0.012$，则 $\theta_\varepsilon=0.03$；如果 $\varepsilon=0.055$，则 $\theta_\varepsilon=0.15$；如果 $\varepsilon=0.138$，则 $\theta_\varepsilon=0.4$。限幅-相关器在低信噪比条件下仍具有稳健性。不过，与线性相关器不同，即使 $\theta>\theta_\varepsilon$，限幅-相关器也不具有一致最优性，因为 d_k'' 取决于θ。实际中得到精确的θ通常较为困难，这限制了相关器和限幅-相关器的工程应用。为此，研究人员已经提出了多种θ未知和θ值较小时的稳健检测方法，具体可参见 Kassam 和 Poor 在 1985 年发表的有关稳健检验的综述性文章。

3.6 习题

1. 请证明，当输入信号为 $\boldsymbol{s}=(s_1, \cdots, s_n)^{\mathrm{T}}$ 且输入噪声均值为 0，协方差为 $\boldsymbol{\Sigma}_N$ 时，滤波器的脉冲响应为

$$\tilde{h}_k = \begin{cases} \tilde{s}_{n-k}, & 0 \leqslant k \leqslant n-1 \\ 0, & 其他 \end{cases}$$

其中 $\tilde{\boldsymbol{s}} = \sum_N^{-1} \boldsymbol{s}$，且在所有的线性滤波器中 n 时刻具有最大输出信噪比。

2. 假设随机观测向量 $\boldsymbol{Y}$ 可表示为

$$Y_k = N_k + \theta S_k, \quad k=1, \cdots, n$$

其中，$\boldsymbol{N}$ 是一个零均值的高斯随机向量，且满足 $E\{N_k N_l\}=\sigma^2\rho^{|k-l|}$，$0\leqslant k, l\leqslant n$，$|\rho|<1$；$\boldsymbol{s}$ 是已知的信号向量。

(a) 请证明下列检验等于在 $\theta=0$ 和 $\theta=1$ 情况下的似然比检验

$$\delta_{(y)} = \begin{cases} 1, & \sum_{k=1}^{n} b_k z_k \geqslant \tau' \\ 0, & \sum_{k=1}^{n} b_k z_k < \tau' \end{cases}$$

其中

$$b_1 = s_1/\sigma$$

$$b_k = (s_k - \rho s_{k-1})/\sigma\sqrt{1-\rho^2}, \quad k = 2, \cdots, n$$

$$z_1 = y_1/\sigma$$

$$z_k = (y_k - \rho y_{k-1})/\sigma\sqrt{1-\rho^2}, \quad k = 2,\cdots,n$$

(b) 根据(a)，求出检测器 ROC 以 ρ/σ、ρ、n 为自变量的函数表达式和虚警概率 α。

3. 考虑 M 进制判定问题($\Gamma=\mathbf{R}^n$)

$$H_0: \boldsymbol{Y} = \boldsymbol{N} + \boldsymbol{s}_0$$

$$H_1: \boldsymbol{Y} = \boldsymbol{N} + \boldsymbol{s}_1$$

$$\vdots$$

$$H_{M-1}: \boldsymbol{Y} = \boldsymbol{N} + \boldsymbol{s}_{M-1}$$

其中，$\boldsymbol{s}_0$，$\boldsymbol{s}_1$，…，$\boldsymbol{s}_{M-1}$是已知信号且具有相同能量$\|\boldsymbol{s}_0\|^2=\|\boldsymbol{s}_1\|^2=\cdots=\|\boldsymbol{s}_{M-1}\|^2$。

(a) 假设 $\boldsymbol{N}\sim\boldsymbol{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$，在所有的假设检验出现概率相等时，请设计实现最小错误概率判定准则。

(b) 进一步假设该信号是正交的，请证明其最小错误概率为

$$P_e = 1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}[\Phi_{(x)}]^{M-1}\mathrm{e}^{-(x-d)^2/2}\mathrm{d}x$$

其中 $d^2=\|\boldsymbol{s}_0\|^2/\sigma^2$。

4. 考虑下面三个关于实数观测序列 Y_1，Y_2，…，Y_n 的三个假设检验

$$H_0: Y_k = N_k - s_k, \quad k = 1,2,\cdots,n$$

$$H_1: Y_k = N_k, \quad k = 1,2,\cdots,n$$

$$H_2: Y_k = N_k + s_k, \quad k = 1,2,\cdots,n$$

其中 s_1，s_2，…，s_n 是一个已知信号序列，N_1，N_2，…，N_n 是服从正态分布$\mathcal{N}(0, 1)$的随机变量序列。

(a) 假设这三个假设检验等概率出现，求出判定其最小化平均错误概率的判决准则。

(b) 假设这三个假设检验等概率出现，计算最小化平均错误概率以判定假设检验。

5. 四相相移键控(QPSK)是习题 3 的一个例子。其中对应的四个信号($M=4$)为

$$s_{lk} = E_0\sin(\omega_c T(k-1) + (l+1/2)\pi/2),$$

$$k = 1,\cdots,n, \quad l = 0,\cdots,3$$

假设参数 ω_c、T、n 和式(3.2.5)一致。求在独立同分布$\mathcal{N}(0, \sigma^2)$噪声下这几个信号等概率出现的最小错误概率(注意：信号相互之间不正交)。

6. 假设 $\boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，其中在 $k\geqslant2$ 时，有$\hat{Y}_k=E\{Y_k|Y_1, \cdots, Y_{k-1}\}$和$\hat{\sigma}^2_{Y_k}=\mathrm{var}(Y_k|Y_1, \cdots, Y_{k-1})$，且$\hat{Y}_1=E(Y_1)$，$\hat{\sigma}^2_{Y_1}=\mathrm{var}(Y_1)$。序列 I_1，I_2，…，I_n 定义如下

$$I_k = (Y_k - \hat{Y}_k)/\hat{\sigma}_{Y_k}$$

证明 $\boldsymbol{I}\sim\mathcal{N}(0, \boldsymbol{I})$，且序列 $\boldsymbol{I}$ 是序列 $\boldsymbol{Y}$ 的白化序列。

7. 考虑下列假设检验对

$$H_0: Y_k = N_k, \quad k = 1,\cdots,n$$

$$H_1: Y_k = N_k + \Theta S_k, \quad k = 1,\cdots,n$$

其中，$\boldsymbol{N}\sim\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$，$\boldsymbol{s}$ 是已知的，Θ 是一个和 $\boldsymbol{N}$ 相互独立的随机变量。

(a) 假设 Θ 是一个以等概率取值+1 和−1 的离散随机变量，求显著水平为 α 的尼曼-皮尔逊检测器和 ROC 曲线。

(b) 假设 $\Theta\sim\mathcal{N}(\boldsymbol{0}, \sigma_\theta^2)$，$\boldsymbol{\Sigma}=\sigma^2\boldsymbol{I}$，证明似然比的公式为

$$L(\boldsymbol{y}) = k_1\mathrm{e}^{k_2\|\boldsymbol{s}^{\mathrm{T}}\boldsymbol{y}\|^2}$$

其中 k_1 和 k_2 是正的常数，并求出 k_2。

8. 假设有观测向量 $Y_k=N_k+\theta S_k$，$k=1$，…，n。其中 $\boldsymbol{N}\sim\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$和 S_1，S_2，…，S_n 是与 $\boldsymbol{N}$ 相互独立的随机变量，且其中每一个都以等概率取值+1 和−1。

(a) 求检验对 $H_0: \theta=0$ 和 $H_1: \theta=A$ 的似然比检验，其中 A 是一个已知常数。

(b) 对假设检验(a)，在 $n=1$ 时，求在虚警概率为 $\alpha\in(0, 1)$下尼曼-皮尔逊准则和相应的检测概率。

(c) 在这个模型中，检验 $H_0: \theta=0$ 和 $H_1: \theta\neq0$ 是否有 UMP 检验？如果有，UMP 是什么？如果没有，原因是什么？(请分别考虑 $n=1$ 和 $n>1$ 两种情况。)

9. 考虑一个 n 维观测随机向量 $\boldsymbol{Y}$，其属于下列两个假设检验中的一个

$$H_0: \boldsymbol{Y} = \boldsymbol{N}$$

$$H_1: \boldsymbol{Y} = \boldsymbol{N} + A[(1-\Theta)\boldsymbol{s}^{(0)} + \theta\boldsymbol{s}^{(1)}]$$

其中 $\boldsymbol{N}\sim\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$；$A$ 是一个非随机的正标量；随机参数 Θ 是一个不依赖于 $\boldsymbol{N}$ 且以等概率取值+1 和−1 的随机变量。信号 $\boldsymbol{s}^{(0)}$ 和 $\boldsymbol{s}^{(1)}$ 是已知的正交向量。即这些信号满足以下条件

$$\sum_{k=1}^{n} s_k^{(l)} s_k^{(m)} = \begin{cases} 1, & m = l \\ 0, & m \neq l \end{cases}$$

(a) 假设 A 的值是已知，求假设检验 H_0 和 H_1 的似然比。

(b) 考虑如下复合假设检验问题

$$H_0: A = 0, \quad H_1: A > 0$$

证明 H_0 和 H_1 的局部最大势检验由下式给出

$$\delta(\boldsymbol{y}) = \begin{cases} 1, & \sum_{k=1}^{n} y_k[s_k^{(0)} + s_k^{(1)}] > \eta \\ \gamma, & \sum_{k=1}^{n} y_k[s_k^{(0)} + s_k^{(1)}] = \eta \\ 0, & \sum_{k=1}^{n} y_k[s_k^{(0)} + s_k^{(1)}] < \eta \end{cases}$$

其中 η 和 γ 是恰当选择的阈值和随机值。

(c) 在(a)的模型中，求(b)检测器的接收机工作特征。

10. 再次考虑习题 9 的模型，但做如下更改
   - 噪声分布是$\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$，其中矩阵 $\boldsymbol{\Sigma}>0$ 不是单位矩阵。
   - 信号 $\boldsymbol{s}^{(0)}$ 和 $\boldsymbol{s}^{(1)}$ 是矩阵 $\boldsymbol{\Sigma}$ 的特征值λ_0 和 λ_1 分别对应的相互正交的特征向量。
   - 参数 Θ 不是随机的，而是等于一个固定值 $\beta\in[0, 1]$。

   分别求出该模型具备最好和最坏检测性能时的 β 值。

11. 推导式(3.2.82)。

12. 差分相移键控(DPSK)是一种二进制的信令方案，它可以对正弦信号进行调制。如果当前调制周期内的载波信号与上一调制周期相同，则传输“0”；如果当前调制周期内的载波信号是上一调制周期内载波信号的 180 度移相，则传输“1”。采用与例 3.2.5 相同的信号和噪声的假设，分析相同代价和等先验概率的贝叶斯判决器性能，并计算错误概率。

13. 考虑如下模型

$$Y_k = \theta^{1/2} s_k R_k + N_k, \quad k = 1, \cdots, n$$

其中 s_1，s_2，…，s_n 是一个已知信号序列，$\theta \geqslant 0$ 是常数，R_1，R_2，…，R_n 和 N_1，N_2，…，N_n 都是服从$\mathcal{N}(0, 1)$的随机变量。

(a) 考虑以下假设检验对

$$H_0: \theta = 0$$
$$H_1: \theta = A$$

其中 A 是一个已知的正常数。请描述尼曼-皮尔逊检测器的结构。

(b) 考虑以下假设检验对

$$H_0: \theta = 0$$
$$H_1: \theta > 0$$

求存在关于 s_1，s_2，…，s_n 的 UMP 检验的条件。

(c) 对于(b)中的假设检验对，其中 s_1，s_2，…，s_n 是更一般的常数，是否存在一个局部最优检测？如果存在，请求解。如果不存在，请描述广义似然比检验。

14. $R_1 = R_2 = \cdots = R_n \sim \mathcal{N}(0, 1)$，且 $\boldsymbol{R}$ 和 $\boldsymbol{N}$ 相互独立，在此条件下重做习题 13。

15. 重新考虑例 3.2.5，其中振幅序列 a_1，a_2，…，a_n 由下式给出

$$a_k = A b_k, \quad k = 1, 2, \cdots, n$$

其中 $\sum_{k=1}^{n} b_k^2 = n$ 和 A 是独立于相位 Θ 的正随机变量，且服从参数为 A_0 的瑞利分布，其概率密度函数为

$$p_{A(a)} = (a/A_0^2)\exp\{-a^2/2A_0^2\}, \quad a \geqslant 0$$

求尼曼-皮尔逊检测器及其门限值 α，并推导出 ROC 的具体表达式。

16. 假设 $p\boldsymbol{s} \sim \mathcal{N}(\mathbf{0}, |\boldsymbol{\Sigma} \boldsymbol{s})$，求解满足下式的 $\hat{\boldsymbol{s}}$

$$\mathrm{e}^{(\hat{\boldsymbol{s}}^{\mathrm{T}} \boldsymbol{y} - 1/2 \| \hat{\boldsymbol{s}} \|^2)/\sigma^2} = \int_{\mathbf{R}^n} \mathrm{e}^{(\boldsymbol{s}^{\mathrm{T}} \boldsymbol{y} - \frac{1}{2} \| \boldsymbol{s} \|^2)/\sigma^2} p\boldsymbol{s}(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}$$

17. 考虑关于实观测值 Y_1，Y_2，…，Y_n 的假设检验

$$H_0: Y_k = N_k, \quad k = 1, 2, \cdots, n$$
$$H_1: Y_k = N_k + \theta s_k, \quad k = 1, 2, \cdots, n$$

其中 N_1，N_2，…，N_n 是服从$\mathcal{N}(0, \sigma^2)$分布的独立随机序列；s_1，s_2，…，s_n 是满足 $\boldsymbol{s}^{\mathrm{T}}\boldsymbol{s}=1$ 的序列；θ 是服从$\mathcal{N}(\mu, v^2)$分布的随机变量，且独立于 N_1，N_2，…，N_n。

(a) 证明尼曼-皮尔逊检验的临界区域为

$$\Gamma_1 = \left\{ \mu \boldsymbol{s}^{\mathrm{T}} \boldsymbol{y} + \frac{v^2}{2\sigma^2} |\boldsymbol{s}^{\mathrm{T}} \boldsymbol{y}|^2 > \tau' \right\}$$

其中 τ' 是阈值。(提示：根据假设 H_1，$\boldsymbol{Y}$ 的协方差矩阵等于 $\sigma^2 I + v^2 \boldsymbol{s}\boldsymbol{s}^{\mathrm{T}}$。)

(b) 求出(a)中所给临界区域对应的虚警和检测概率。

18. 从式(3.2.111)推导出式(3.2.112)。

19. 用 $\mu_{T,0}(s)$ 表示 H_0 假设下对数似然比的累积量母函数，并假设 $\mu_{T,0}(s)$ 二次可微。

(a) 证明 $(\mu_{T,0}(s)-s\tau)$ 是 s 的凸函数。

(b) 证明

$$\mu'_{T,0}(j)=E\{\log L(Y)\mid H_j\},\quad j=0,1$$

其中 $\mu'_{T,0}(s)=\frac{\mathrm{d}}{\mathrm{d}s}\mu_{T,0}(s)$。

(c) 证明如果 $\min\limits_{s\geqslant 0}[\mu_{T,0}(s)-s\tau]$ 在 $s\geqslant 1$ 时取最小值，那么在 $s=1$ 时 $\min\limits_{s\leqslant 1}[\mu_{T,0}(s)+(1-s)\tau]$ 取最小值且最小值为零。

(d) 证明如果 $\min\limits_{s\leqslant 1}[\mu_{T,0}(s)-s\tau]$ 在 $s\leqslant 0$ 时取最小值，那么在 $s=0$ 时 $\min\limits_{s\geqslant 0}[\mu_{T,0}(s)-s\tau]$ 取最小值且最小值为零。

20. 推导式(3.3.15)。

21. 计算二进制对称信道在等于先验概率 $\pi_0=\pi_1=1/2$ 下的切诺夫界，并将它与实际最小错误概率进行比较。

22. 考虑如下假设检验对

$$H_0: Y_k=N_k-s_k,\quad k=1,\cdots,n$$
$$H_1: Y_k=N_k+s_k,\quad k=1,\cdots,n$$

其中 N_1，N_2，…，N_n 是独立同分布的拉普拉斯随机变量，s_1，s_2，…，s_n 是满足 $s_k\geqslant\Delta>0$ 的已知信号。证明：当 $n\to\infty$ 时，最小错误概率趋近于零。

23. 考虑在 $\mathcal{N}(\mathbf{0},\sigma^2\boldsymbol{I})$ 噪声中检测服从 $\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{\mathrm{S}})$ 分布的信号，其中 $n=2$ 且有

$$\boldsymbol{\Sigma}_{\mathrm{S}}=\sigma_s^2\begin{pmatrix}1&\rho\\\rho&1\end{pmatrix}$$

对于相等的先验概率，分别计算 $\rho=0.0$、$\rho=-0.5$ 和 $\rho=+0.5$，$\sigma_s^2/\sigma^2=0.1$、$\sigma_s^2/\sigma^2=1.0$ 和 $\sigma_s^2/\sigma^2=10.0$ 时错误概率和切诺夫界。

24. 对于独立同分布的观测序列 Y_1，Y_2，…，Y_n，计算下列边缘概率密度函数对应的切诺夫界

$$p_0(y)=\begin{cases}1, & 0\leqslant y\leqslant 1\\0, & \text{其他}\end{cases}$$

$$p_1(y)=\begin{cases}2y, & 0\leqslant y\leqslant 1\\0, & \text{其他}\end{cases}$$

25. 独立同分布的伯努利观测序列 Y_1，Y_2，…的分布为

$P(Y_k=1)=1-P(Y_k=0)=1/3$，　在 H_0 下

和

$P(Y_k=1)=1-P(Y_k=0)=2/3$，　在 H_1 下

(a) 利用 Wald 近似计算值 A 和 B，使得 SPRT(A，B) 的最大错误概率 $p^*=\max(p_{\mathrm{F}}, p_{\mathrm{M}})$ 约等于 0.01。采用 Wald 逼近，求解预期的样本大小 $E\{N\mid H_0\}$ 和 $E\{N\mid H_1\}$。

(b) 求解尽可能小的整数 n，使得固定样本数为 n 的最优判决的最大误差概率不超过 0.01。(提示：可根据切诺夫界计算 n。)

(c) 精确计算(a)中的 p^*、$E\{N\mid H_0\}$ 和 $E\{N\mid H_1\}$，并与(a)中的近似值进行比较。(提示：在(a)中使用 SPRT 得到的值相当于 SPRT(A'，B')，其中 A' 和 B' 是 2 的整数次幂。)

26. 假设 N_1，N_2，…，N_n 是均值为 0、方差为 1 的独立高斯随机变量，S_1，S_2，…，S_n 是均值为 0、方差为 3 的独立高斯随机变量，且相互独立，考虑如下假设检验对

$$H_0: Y_k=N_k,\qquad k=1,2,\cdots$$
$$H_1: Y_k=S_k+N_k,\quad k=1,2,\cdots$$

(a) 在该模型下重做习题 25(a)。

(b) 在该模型下重做习题 25(b)。

27. 推导式(3.5.22)。

28. 证明式(3.5.24)和式(3.5.25)是相同的。

29. 推导式(3.5.26)。

30. 验证式(3.5.54)。

第4章

参数估计基础

4.1 引言

第 2 章和第 3 章讨论了如何在随机观测 Y 的基础上从两个可能的统计假设中设计最优判决准则。在许多实际应用中，感兴趣的不是在两个(或多个)离散情形中作选择，而是在可能的连续状态中做出选择。特别地，在第 2 章讨论过的复合假设检验问题中，可以考虑将观测空间的一类分布由一个或者一组参数标识。但是不同于复合假设检验中期望得到关于参数的二元决策，这里期望尽可能精确地从观测数据中确定参数的实际值。

这类问题称为参数(点)估计问题，本章将讨论与参数估计最优算法有关的基础理论。在假设检验问题(也可认为是参数估计问题的一种特殊情况)中，可以使用各种估计原理，这些原理的不同之处主要在于参数已知先验信息的多少和所使用的性能评价标准。

本章讨论两种基本的参数估计方法——一种是贝叶斯方法，该方法将参数看成一个与观测统计相关的随机量；另一种方法把参数看成一个未被赋予任何概率结构的未知量。在两种方法中，贝叶斯方法是最简单易懂的，因此在 4.2 节中首先予以讨论，非随机参数的估计将会在本章其余章节中进行介绍。

需要注意的是，本章只讨论静态参数估计问题，即待估计的参数不随时间变化。估计随时间变化的参数(如信号)将在第 5 章中讨论。

4.2 贝叶斯参数估计

在本章中，假设一个随机观测 Y 的分布族模型可由参数 θ 标识，其中 θ 值组成的集合为 Λ。比如，得到一个集合 $\{P_\theta;\ \theta\in\Lambda\}$，其中 P_θ 表示观测空间 $(\Gamma, \mathcal{G})$ 中的一个分布。同时，假设参数集 Λ 为 m 维空间 $\mathbf{R}^m$ 的子空间。在这个模型中，参数估计问题的目标就是找到一个函数 $\hat{\theta}$：$\Gamma\to\Lambda$，使得 $\hat{\theta}(y)$ 是在观测 $Y=y$ 的基础上最接近 θ 的真实值(例如，$Y\sim P_\theta$ 时的 θ 值)。

当然，参数估计问题的解取决于衡量估计性能的标准，所以在假设检验问题中，首先要做的是分配参数决策的代价。特别地，假设有一个函数 C：$\Lambda\times\Lambda\to\mathbf{R}$ 使得 $C[a, \theta]$ 是将 θ 估计成 a 的代价，其中 a 和 θ 均包含在 Λ 中。一旦给定函数 C 之后，对于每个 $\theta\in\Lambda$，可以把 Y 的条件风险或平均代价与估计值 $\hat{\theta}$ 联系起来，即

$$R_\theta(\hat{\theta}) = E_\theta\{C[\hat{\theta}(Y),\theta]\} \tag{4.2.1}$$

如果把参数的实际值 θ 看成是随机变量 Θ 的某次实现值，那么可以定义平均风险或贝叶斯风险为

$$r(\hat{\theta}) \triangleq E\{R_\Theta(\hat{\theta})\} \tag{4.2.2}$$

合理的设计目标是找到一个估计值使得 $r(\hat{\theta})$ 最小，对应的估计值就是 θ 的贝叶斯估计。

注意，$R_\theta(\hat{\theta})=E\{C[\hat{\theta}(Y), \Theta]|\Theta=\theta\}$，由此可以得到

$$r(\hat{\theta}) = E\{C[\hat{\theta}(Y),\Theta]\} = E\{E\{C[\hat{\theta}(Y),\Theta]|Y\}\} \tag{4.2.3}$$

由式(4.2.3)可以看出，θ 的贝叶斯估计(如果存在的话)可以通过对每个 $y\in\Gamma$ 计算给定 $Y=y$ 条件下的最小化后验代价得到

$$E\{C[\hat{\theta}(y),\Theta]|Y = y\} \tag{4.2.4}$$

这和贝叶斯假设检验问题的求解步骤相同(见 2.5 节)。需要注意的是：如果假设 Θ 在给定 $Y=y$ 时具有条件概率密度 $w(\theta|y)$，则 $y\in\Gamma$ 的贝叶斯估计 $\hat{\theta}(y)$ 可以通过求以下积分的极小值得到

$$\int_\Lambda C[\hat{\theta}(y),\theta]w(\theta|y)\mu(\mathrm{d}\theta) \tag{4.2.5}$$

下面通过不同的模型来说明贝叶斯准则的具体应用。

模型 4.2.1　最小均方误差(MMSE)估计

当 $\Lambda=\mathbf{R}$ 且 $E\{\Theta^2\}<\infty$ 时，一个应用广泛的代价函数为

$$C[a,\theta] = (a-\theta)^2, \quad (a,\theta) \in \mathbf{R}^2 \tag{4.2.6}$$

这个代价函数是很多情况下的首选，因为它以估计误差 $\hat{\theta}(y)-\theta$ 平方的形式来衡量估计器的性能。根据上述代价函数，对应的贝叶斯风险为 $E\{[\hat{\theta}(Y)-\Theta]^2\}$，也称为均方误差(MSE)。此时，贝叶斯估计与最小均方误差(MMSE)估计等价。

在本例中，给定 $Y=y$ 的后验代价为

$$\begin{aligned}E\{[\hat{\theta}(y)-\Theta]^2|Y = y\} &= E\{[\hat{\theta}(y)]^2|Y = y\} - 2E\{\hat{\theta}(y)\Theta|Y = y\} + E\{\Theta^2|Y = y\} \\ &= [\hat{\theta}(y)]^2 - 2\hat{\theta}(y)E\{\Theta|Y = y\} + E\{\Theta^2|Y = y\}\end{aligned} \tag{4.2.7}$$

式(4.2.7)是关于 $\hat{\theta}(y)$ 的二次函数，在导数等于零处取得唯一最小值。为了与式(4.2.7)相区分，用 $\hat{\theta}_{\mathrm{MMSE}}(y)$ 表示贝叶斯估计

$$\hat{\theta}_{\mathrm{MMSE}}(y) = E\{\Theta|Y = y\} \tag{4.2.8}$$

因此，给定 $Y=y$ 时对 Θ 的 MMSE 估计就是给定 $Y=y$ 时 Θ 的条件均值。这是一个非常基本的结论，在随后的章节中将用到此结论。在某些情况下，MMSE 估计也称为条件均值估计(CME)。

模型 4.2.2　最小绝对值误差估计(MMAE)

对于 $\Lambda=\mathbf{R}$ 的情况，绝对误差是另一个可以应用的代价函数，其表达式为

$$C[a,\theta] = |a-\theta|, \quad (a,\theta) \in \mathbf{R}^2 \tag{4.2.9}$$

其贝叶斯风险为 $E\{|\hat{\theta}(Y)-\Theta|\}$，称为绝对均值误差，因此对应的贝叶斯估计称为最小绝对值误差(MMAE)估计。

为了推导出 MMAE 估计，需要利用下述结论：如果 X 是一个概率分布为 $P(X\geqslant 0)=1$ 的随机变量，那么 $E\{X\} = \int_0^\infty P(X > x)\mathrm{d}x$。该结论本质上可由分部积分得到[参考 Breiman (1968)]。

由于$|\hat{\theta}(y)-\Theta|\geqslant 0$，运用以上结论可以得到

$$E\{|\hat{\theta}(y)-\Theta\|Y=y\}=\int_0^{\infty}P(|\hat{\theta}(y)-\Theta|>x|Y=y)\mathrm{d}x=\int_0^{\infty}P(\Theta>x+\hat{\theta}(y)|Y=y)\mathrm{d}x$$
$$+\int_0^{\infty}P(\Theta<-x+\hat{\theta}(y)|Y=y)\mathrm{d}x \tag{4.2.10}$$

式(4.2.10)右边第一个积分中令$t=x+\hat{\theta}(y)$，第二个积分中令$t=-x+\hat{\theta}(y)$，则有

$$E\{|\hat{\theta}(y)-\Theta\|Y=y\}=\int_{\hat{\theta}(y)}^{\infty}P(\Theta>t|Y=y)\mathrm{d}t+\int_{-\infty}^{\hat{\theta}(y)}P(\Theta<t|Y=y)\mathrm{d}t \tag{4.2.11}$$

式(4.2.11)中的$E\{|\hat{\theta}(y)-\Theta\|Y=y\}$是一个关于$\hat{\theta}(y)$的微分函数。两边同时微分可得

$$\frac{\partial}{\partial\hat{\theta}(y)}E\{|\hat{\theta}(y)-\Theta\|Y=y\}=P(\Theta<\hat{\theta}(y)|Y=y)-P(\Theta>\hat{\theta}(y)|Y=y) \tag{4.2.12}$$

从式(4.2.12)可以看出，此导数是一个关于$\hat{\theta}(y)$的单调非减函数，当$\hat{\theta}(y)\to-\infty$时趋近于$-1$；当$\hat{\theta}(y)\to+\infty$时趋近于$+1$。因此$E\{|\hat{\theta}(y)-\Theta\|Y=y\}$在其导数变号时获得最小值点(或最小值点集)。在这种情况下，贝叶斯估计[记为$\hat{\theta}_{\mathrm{ABS}}(y)$]是满足下列条件的任意点。

$$\begin{cases}P(\Theta<t|Y=y)\leqslant P(\Theta>t|Y=y), & t>\hat{\theta}_{\mathrm{ABS}}(y)\\ P(\Theta<t|Y=y)\geqslant P(\Theta>t|Y=y), & t<\hat{\theta}_{\mathrm{ABS}}(y)\end{cases} \tag{4.2.13}$$

注意，满足式(4.2.13)的$\hat{\theta}_{\mathrm{ABS}}(y)$是给定$Y=y$时$\Theta$的条件分布的中位数。因此 MMAE 估计是一种条件中位数估计(CME)。当且仅当给定$Y=y$时Θ的条件分布的平均数和中位数相等时，条件中位数估计才和 MMAE 估计一致。当然，这两种关于Θ的估计哪一种更好需要取决于所采用的标准。

模型 4.2.3　最大后验概率(MAP)估计

最大后验概率(MAP)估计虽然不完全是贝叶斯估计，但是它适合在贝叶斯模式下使用。

为了推导这种方法，假设$\Lambda=\mathbf{R}$情形，考虑均匀代价函数

$$C[a,\theta]=\begin{cases}0, & |a-\theta|\leqslant\Delta\\ 1, & |a-\theta|>\Delta\end{cases} \tag{4.2.14}$$

其中$\Delta>0$。此时，对于估计量$\hat{\theta}$，给定$Y=y$时的平均后验代价为

$$E\{C[\hat{\theta}(y),\Theta]|Y=y\}=P(|\hat{\theta}(y)-\Theta|>\Delta|Y=y)$$
$$=1-P(|\hat{\theta}(y)-\Theta|\leqslant\Delta|Y=y) \tag{4.2.15}$$

为了得到式(4.2.15)的最小值，首先假设Θ是一个离散随机变量，值取自有限集合$\Lambda=\{\theta_0,\cdots,\theta_{M-1}\}$，当$i\neq j$时，$|\theta_i-\theta_j|>\Delta$，可以得到

$$E\{C[\hat{\theta}(y),\Theta]|Y=y\}=1-P(\Theta=\hat{\theta}(y)|Y=y)$$
$$=1-w(\hat{\theta}(y)|y),\quad \hat{\theta}(y)\in\Lambda \tag{4.2.16}$$

其中$w(\theta|y)$为给定$Y=y$时Θ的条件概率密度函数。可以从式(4.2.16)中看出，此时的贝叶斯估计是由在$y\in\Gamma$、$\theta\in\Lambda$中使得$w(\theta|y)$取最大值的任意θ值得到。即贝叶斯估计是当$Y=y$时有最大后验概率的Θ值。[⊖]

⊖ 当然，当Λ是一个有限集时，这种情况是一个简单的M元假设检验问题，因为$i\neq j$时有$|\theta_i-\theta_j|>\Delta$，所以式(4.2.14)中的代价函数退化为：若$a\neq\theta$，则$C[a,\theta]=1$；若$a=\theta$，则$C[a,\theta]=0$。在这种情况下，贝叶斯估计变为对于均匀代价的$M$元贝叶斯决策规则(见第 2 章习题 16)。

假设 $\Lambda=\mathbf{R}$ 且 Θ 是一个连续随机变量，给定 $Y=y$ 时的条件概率密度函数为 $w(\theta|y)$。在这种情况下后验风险变为

$$E\{C[\hat{\theta}(y),\Theta]|Y=y\}=1-\int_{\hat{\theta}(y)-\Delta}^{\hat{\theta}(y)+\Delta}w(\theta|y)\mathrm{d}\theta \tag{4.2.17}$$

当 $w(\theta|y)$ 在区间 $(\hat{\theta}(y)-\Delta,\ \hat{\theta}(y)+\Delta)$ 取关于 $\hat{\theta}(y)$ 的最大值时，式(4.2.17)有最小值。由图 4.2.1 可以看出，如果 $w(\theta|y)$ 是关于 θ 的平滑函数且 Δ 足够小，那么可以通过选择 $\hat{\theta}(y)$ 为 $w(\theta|y)$ 的最大值点得到这一区域的近似最大化，进而得到式(4.2.17)的最小值。换言之，对于足够小的 Δ 和连续的 $w(\theta|y)$，有下面的表达式成立

$$\int_{\hat{\theta}(y)-\Delta}^{\hat{\theta}(y)+\Delta}w(\theta|y)\mathrm{d}\theta\approx 2\Delta w(\theta|y)\big|_{\theta=\hat{\theta}(y)} \tag{4.2.18}$$

在 Λ 中选择 $\hat{\theta}(y)$ 作为 θ 代入 $w(\theta|y)$ 得到最大值，从而使等式右边也得到最大值。

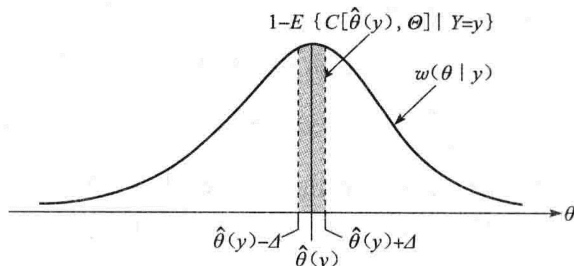


图 4.2.1 MAP 估计

在以上几个情况中，均匀代价准则使得估计 Θ 的过程变为求解后验(离散或连续)密度 $w(\theta|y)$ 的最大值的过程。[类似地，当 θ 在无穷离散序列中取值时，可以证明，式(4.2.15)通过选择 $\hat{\theta}(y)$ 得到条件密度函数 $w(\theta|y)$ 的最大值的方法来找到最小值。]这种估计称为最大后验概率估计(MAP)，记为 $\hat{\theta}_{\mathrm{MAP}}$。虽然 MAP 估计只是 Δ 足够小情况下以均匀代价函数为准则的贝叶斯估计的近似，但是 MAP 准则在估计器的设计中运用广泛，其中的一个关键因素是它通常比 MMSE、MMAE 或其他估计方法更容易计算。

需要注意的是，密度取得最大值的点称为相应概率分布的众数。因为 $\hat{\theta}_{\mathrm{MAP}}$ 是通过其条件分布的众数来估计 Θ，所以称为条件众数估计。

从模型 4.2.1～模型 4.2.3(和模型 4.2.5)可以看出，给定条件的贝叶斯估计取决于给定观测中参数的条件分布。特别地，MMSE 估计、MMAE 估计和 MAP 估计分别为分布的平均值、中位数和众数。就假设检验而言，可以将观测理解为一种将参数的先验分布转换为后验分布的过程。一般地，贝叶斯估计具有后验分布的特征。

在对统计过程进行建模时，一般首先从给定 $\Theta=\theta$ 时 Y 的条件分布族 $\{P_\theta;\ \theta\in\Lambda\}$ 开始，对于贝叶斯公式也会有一个关于 Θ 的先验分布。从给定 Y 时 Θ 的先验和条件分布中得到给定 Y 时关于 Θ 的条件分布时，只需要用到贝叶斯公式。特别地，假设对于每一个 $\theta\in\Lambda$，P_θ 都有一个密度 p_θ，且 Θ 的先验分布密度函数为 $w(\theta)$，则给定 $Y=y$ 时关于 Θ 的条件分布的概率密度为

$$w(\theta|y)=\frac{p_\theta(y)w(\theta)}{\int_\Lambda p_\theta(y)w(\theta)\mu(\mathrm{d}\theta)} \tag{4.2.19}$$

注意，式(4.2.19)的分母是 $p(y)$，它是 Y 的非条件概率密度[⊖]。

以上三种情况的贝叶斯估计均可以从式(4.2.19)直接推出。不过，MAP 估计可以不

⊖ 也称为全概率。——译者注

用计算 $p(y)$，因为该项不会影响 θ 的最大值。也就是说，$\hat{\theta}_{\mathrm{MAP}}(y)$ 是通过在 $\theta\in\Lambda$ 中求解 $p_\theta(y)w(\theta)$ 关于 θ 的最大值而得到。考虑到对数函数为单调增函数，对于 $\theta\in\Lambda$，$\hat{\theta}_{\mathrm{MAP}}(y)$ 同样是 $(\log p_\theta(y)+\log w(\theta))$ 的最大值。如果 Θ 是一个给定 $Y=y$ 时的连续随机变量，并且 p_θ 和 w 足够光滑，则取得最大值的必要条件是

$$\frac{\partial}{\partial\theta}\log p_\theta(y)\big|_{\theta=\hat{\theta}_{\mathrm{MAP}}(y)}=-\frac{\partial}{\partial\theta}\log w(\theta)\big|_{\theta=\hat{\theta}_{\mathrm{MAP}}(y)} \tag{4.2.20}$$

式(4.2.20)称为 MAP 方程。

下面的两个例子用于说明 MMSE 估计、MMAE 估计和 MAP 估计的计算。

例 4.2.1 指数分布的参数估计

考虑 $\Lambda=(0,\infty)$ 且 $\Gamma=\mathbf{R}$，给定 $\Theta=\theta$ 时观测数据的条件概率密度函数如下

$$p_\theta(y)=\begin{cases}\theta\mathrm{e}^{-\theta y}, & y\geqslant 0\\ y, & y<0\end{cases} \tag{4.2.21}$$

这就是参数为 θ 的指数密度。指数密度可以对许多物理现象进行建模，特别是在建模随机时间发生的连续事件的时间间隔方面非常有效，如到达通信转接站的信息或数据包、到达交叉路口的车辆、从相干光源中发射出的光子或逻辑电路中出现的器件故障等。模型中的参数 θ 可以解释为这些事件发生的频率，因此可以把此处的估计问题看成从这些事件连续发生的时间观测来估计它们发生的概率。

假设 Θ 的先验信息也服从指数分布，其分布密度为

$$w(\theta)=\begin{cases}\alpha\mathrm{e}^{-\alpha\theta}, & \theta\geqslant 0\\ 0, & \theta<0\end{cases} \tag{4.2.22}$$

其中已知 $\alpha>0$。从式(4.2.19)中可以得到给定 $Y=y$ 时 Θ 的后验分布为

$$w(\theta|y)=\frac{\alpha\theta\mathrm{e}^{-(\alpha+y)\theta}}{\int_0^\infty\alpha\theta\mathrm{e}^{-(\alpha+y)\theta}\mathrm{d}\theta}=(\alpha+y)^2\theta\mathrm{e}^{-\theta(\alpha+y)} \tag{4.2.23}$$

其中 $\theta\geqslant0$ 且 $y\geqslant0$，其他情况时 $w(\theta|y)=0$。

MMSE 估计就是式(4.2.23)的均值，由下式给出

$$\hat{\theta}_{\mathrm{MMSE}}(y)=\int_0^\infty\theta w(\theta|y)\mathrm{d}\theta=(\alpha+y)^2\int_0^\infty\theta^2\mathrm{e}^{-\theta(\alpha+y)}\mathrm{d}\theta=\frac{2}{\alpha+y} \tag{4.2.24}$$

注意，对于固定的 α，Θ 的估计反比于 y。从前面对指数模型的说明可以看出这个结果是直观合理的，大的时间间隔(较大的 y)就意味着低的发生频率(较小的 θ)。因为 MMSE 的估计值不大于 $2/\alpha$，所以在一定程度上估计性能取决于 α 值。α 的值很小表明 Θ 的分布很广泛[即 $w(\theta)$ 相对发散]，因此通过观测数据得到的估计量也允许 Θ 有较大误差。与之相反，较大的 α 值表明 Θ 接近 0 的可能性较大，而对应的估计均方误差较小。

在这种情况下，MSE 的最小值可以直接计算出来。首先从式(4.2.3)可以得到贝叶斯风险是后验代价的平均值，因而有

$$\begin{aligned}\mathrm{MMSE}&=r(\hat{\theta}_{\mathrm{MMSE}})=E\{E\{(\hat{\theta}_{\mathrm{MMSE}}(Y)-\Theta)^2|Y\}\}\\&=E\{E\{[\Theta-E\{\Theta|Y\}]^2|Y\}\}=E\{\mathrm{var}(\Theta|Y)\}\end{aligned} \tag{4.2.25}$$

因此，MSE 的最小值是给定 Y 时 Θ 的条件方差的平均值。因为

$$\mathrm{var}(\Theta|Y=y)=E\{\Theta^2|Y=y\}-E^2\{\Theta|Y=y\},$$

所以

$$\mathrm{var}(\Theta|Y=y)=\int_0^\infty\theta^2w(\theta|y)\mathrm{d}\theta-[\hat{\theta}_{\mathrm{MMSE}}(y)]^2=(\alpha+y)^2\int_0^\infty\theta^3\mathrm{e}^{-\theta(\alpha+y)}\mathrm{d}\theta-\frac{4}{(\alpha+y)^2}=\frac{2}{(\alpha+y)^2}$$

进而有

$$\mathrm{MMSE}=E\left\{\frac{2}{(\alpha+Y)^2}\right\}=\int_0^\infty\frac{2}{(\alpha+y)^2}p(y)\mathrm{d}y=\int_0^\infty\frac{2\alpha}{(\alpha+y)^4}\mathrm{d}y=\frac{2}{3\alpha^2}\tag{4.2.26}$$

这里运用了式(4.2.23)中的 $p(y)=\int_0^\infty\alpha\theta\mathrm{e}^{-(\alpha+y)\theta}\mathrm{d}\theta=\alpha/(\alpha+y)^2$ 。

MMAE 估计$\hat{\theta}_{\mathrm{ABS}}(y)$是 $w(\theta|y)$的中位数。由于给定 $Y=y$ 时 Θ 是连续的，$\hat{\theta}_{\mathrm{ABS}}(y)$可以通过求解下面的方程得到

$$\int_{\hat{\theta}_{\mathrm{ABS}}(y)}^\infty w(\theta|y)\mathrm{d}\theta=\frac{1}{2}\tag{4.2.27}$$

代入式(4.2.23)并进行积分后，有下面的表达式成立

$$[1+(\alpha+y)\hat{\theta}_{\mathrm{ABS}}(y)]\mathrm{e}^{-(\alpha+y)\hat{\theta}_{\mathrm{ABS}}(y)}=\frac{1}{2}\tag{4.2.28}$$

进而可求出

$$\hat{\theta}_{\mathrm{ABS}}(y)=\frac{T_o}{\alpha+y}\tag{4.2.29}$$

其中 T_o 是式$(1+T_o)\mathrm{e}^{-T_o}=1/2$ 的解，近似解为 $T_o\approx1.68$。比较式(4.2.29)和式(4.2.24)可以发现，MMAE 和 MMSE 具有相似的估计性能，这些差异仅仅在于分子的常数不同。在这种情况下，求解最小贝叶斯风险可以类似地用计算 MMSE 估计最小值的方法，具体计算过程作为章后练习。

在这种情况下，Θ 的 MAP 估计也可以很容易地计算出来。注意

$$\frac{\partial}{\partial\theta}[\log p_\theta(y)+\log w(\theta)]=\frac{\partial}{\partial\theta}(\log\theta-\theta y+\log\alpha-\alpha\theta)=\theta^{-1}-(\alpha+y)$$

并且

$$\frac{\partial^2}{\partial\theta^2}[\log p_\theta(y)+\log w(\theta)]=-\theta^{-2}<0$$

可以看出 $w(\theta|y)$有唯一的最大值，取得最大值的点为

$$\hat{\theta}_{\mathrm{MAP}}(y)=\frac{1}{\alpha+y}\tag{4.2.30}$$

显然，再次得到一个与 MMSE 仅相差一个尺度系数的估计结果。◀

在例 4.2.1 中，根据三个不同的估计准则得到了三个不同的关于 Θ 的估计量。为了决定使用哪种估计器，必须结合特定应用环境，在三个相应的代价函数中确定哪一个对应用所涉及的估计误差的算法更合适。在很多应用问题中，因为这三种估计是一致的，所以一般不需选择。下面就是此种情况的一个例子。

例 4.2.2 信号幅度的估计

考虑如下情况：$\Gamma=\mathbf{R}^n$，$\Lambda=\mathbf{R}$，并且

$$Y_k=N_k+\Theta s_k\quad k=1,\cdots,n\tag{4.2.31}$$

其中 $\boldsymbol{N}\sim\mathcal{N}(0,\ \boldsymbol{\Sigma})$，$s$ 已知，$\Theta\sim\mathcal{N}(\mu,\ v^2)$，$\boldsymbol{N}$ 和 Θ 是相互独立的。事实上，该问题与加性噪声下估计确知信号的未知幅度问题相对应。

给定 $\Theta=0$，得到 $\boldsymbol{Y}\sim\mathcal{N}(\theta\boldsymbol{s},\ \boldsymbol{\Sigma})$。因此，$\Theta$ 的后验密度为

$$\omega(\theta|\boldsymbol{y})=\frac{\dfrac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}\mathrm{e}^{-(1/2)(\boldsymbol{y}-\theta\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\theta\boldsymbol{s})}\dfrac{1}{\sqrt{2\pi}v}\mathrm{e}^{-(\theta-\mu)^2/2v^2}}{\displaystyle\int_{-\infty}^{\infty}\frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}}\mathrm{e}^{-(1/2)(\boldsymbol{y}-\theta\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\theta\boldsymbol{s})}\frac{1}{\sqrt{2\pi}v}\mathrm{e}^{-(\theta-\mu)^2/2v^2}\mathrm{d}\theta}$$

$$=K(\boldsymbol{y})\exp\left\{-\frac{\theta^2}{2}(d^2+1/v^2)+\theta\left(\boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}+\frac{\mu}{v^2}\right)\right\}\tag{4.2.32}$$

正如在第 3 章中定义的 $d^2=\boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{s}$ 一样，式中的 $K(\boldsymbol{y})$ 是一个依赖于 $\boldsymbol{y}$，但与 θ 无关的函数。注意，式(4.2.32)中的 $w(\theta|\boldsymbol{y})$ 是 θ 二次项的指数函数，因此它必然服从高斯分布。如果 $w(\theta|\boldsymbol{y})$ 服从 $\mathcal{N}(m,\ q^2)$，可以得到

$$w(\theta|\boldsymbol{y})=\frac{1}{\sqrt{2\pi}q}\mathrm{e}^{-(\theta-m)^2/2q^2}=\frac{\mathrm{e}^{-m^2/2q^2}}{\sqrt{2\pi}q}\mathrm{e}^{-\theta^2/2q^2+\theta m/q^2}\tag{4.2.33}$$

将式(4.2.32)和式(4.2.33)进行比较，可见在给定 $\boldsymbol{Y}=\boldsymbol{y}$、$\Theta\sim\mathcal{N}(m,\ q^2)$ 条件下，有

$$q^2=(d^2+1/v^2)^{-1}$$

并且

$$m=(d^2+1/v^2)^{-1}(\boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}+\mu/v^2)$$

此时 $K(\boldsymbol{y})$ 的值为 $\mathrm{e}^{-m^2/2q^2}/\sqrt{2\pi}q$。

因为高斯分布的第一个参数是其均值，所以立即能得到 Θ 的条件均值估计为

$$\hat{\theta}_{\mathrm{MMSE}}(\boldsymbol{y})=\frac{\boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}+\mu/v^2}{d^2+1/v^2}=\frac{v^2d^2\ \hat{\theta}_1(\boldsymbol{y})+\mu}{v^2d^2+1}\tag{4.2.34}$$

其中 $\hat{\theta}_1(y)\triangleq\boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{y}/d^2$。此外，由于 $\mathrm{var}(\Theta|\boldsymbol{Y})=(d^2+1/v^2)^{-1}$ 不依赖于 $\boldsymbol{Y}$，因而最小均方误差为

$$\mathrm{MMSE}=E\{\mathrm{var}(\Theta|\boldsymbol{Y})\}=\frac{1}{d^2+1/v^2}=\frac{v^2}{v^2d^2+1}\tag{4.2.35}$$

同时，由于高斯分布关于其均值对称且在均值处达到最大值，条件中位数和条件众数都与条件平均数相等，因此有如下关系：$\hat{\theta}_{\mathrm{ABS}}=\hat{\theta}_{\mathrm{MAP}}=\hat{\theta}_{\mathrm{MMSE}}$。

本例能充分体现贝叶斯估计的一些特性。注意，v^2 决定了关于 Θ 的先验知识的准确度，即 v^2 越小，在缺乏观测的情况下所了解的 θ 越准确。另一方面，从第 3 章高斯噪声中相干信号检测的有关论述来看，d^2 是一种量度，它可以衡量 $\boldsymbol{s}$ 与服从 $\mathcal{N}(0,\ \boldsymbol{\Sigma})$ 分布的噪声的分离程度。换言之，就携带信息的信号而言，d^2 是观测精确度的量度——在这个意义上，大的 d^2 值与高质量的观测相对应，小的 d^2 值与低质量的观测相对应。

理解这些概念之后，重新考虑式(4.2.34)中的估计值 $\hat{\theta}_{\mathrm{MMSE}}$。如果 v^2d^2 与其他量相比非常小，那么就可以得到 $\hat{\theta}_{\mathrm{MMSE}}(\boldsymbol{y})\approx\mu$。这发生在先验知识与观测相比非常精确的时候(也就是说，v^2 相对小于 $1/d^2$)，估计器忽略观测而选择先验分布的均值作为其估计值。此时，最小均方误差约等于 v^2，即先验方差。另一方面，如果 v^2d^2 的值较大，那么 $\hat{\theta}_{\mathrm{MMSE}}(\boldsymbol{y})\approx\hat{\theta}_1(\boldsymbol{y})$，即估计值只取决于观测而与先验信息无关。该估计值同样是合理的，这

是因为当v^2远大于$1/d^2$时，相比于先验信息而言，最好还是相信观测的结果。后一种情形下的最小均方误差约等于$1/d^2$[⊖]。在这两种极端情况之间，最优估计器要平衡先验知识和观测结果，相应的最小均方误差则体现了这种平衡。

接着上面的讨论，考虑$\boldsymbol{\Sigma}=\sigma^2 I$和$\mathbf{s}=\mathbf{1}\triangleq(1,\ 1,\ \cdots,\ 1)^{\mathrm{T}}$这一特殊情况。在这种情况下的观测是

$$Y_k = N_k + \Theta,\quad k = 1,\cdots,n$$

其中N_1，…，N_n服从$\mathcal{N}(0,\ \sigma^2)$。量值$v^2 d^2 = nv^2/\sigma^2$，样本均值$\hat{\theta}_1(\mathbf{y}) = \bar{y} \triangleq (1/n)\sum_{k=1}^{n} y_k$。如果缺乏观测数据($n=0$)，可以简单地把$\Theta$估计成其先验均值$\mu$，但是能得到更多的观测值(增加$n$)，样本均值$\bar{y}$就会变得更加可靠，进而可对其赋予更大的权值。在$n\to\infty$时，可以完全忽略先验均值，而采用样本均值作为最终估计。先验信息和样本均值的加权比例受到比率v^2/σ^2的控制(σ^2决定了每一次观测的精确度)。

前面的讨论将注意力集中在了估计单个实参数问题上。然而，在实践过程中遇到的许多问题要求同时估计多个参数。当然，贝叶斯公式在估计向量参数时同样适用，在下面的讨论中将研究这一问题。◀

模型 4.2.4　向量参数估计

现在考虑$\Lambda=\mathbf{R}^m$的情形。为了按照贝叶斯步骤设计$\boldsymbol{\Theta}$的估计器，必须指定一个代价函数C：$\mathbf{R}^m\times\mathbf{R}^m\to\mathbf{R}$。在一些情况下，可使用如下形式的代价函数

$$C[\boldsymbol{a},\boldsymbol{\theta}] = \sum_{i=1}^{m} C_i[a_i,\theta_i] \tag{4.2.36}$$

其中C_i是与估计参数的第i个分量有关的代价函数。确定代价函数后，则估计值$\hat{\boldsymbol{\theta}}$的条件后验代价按照下式计算

$$E\{C[\hat{\boldsymbol{\theta}}(y),\boldsymbol{\Theta}]\,|\,Y = y\} = \sum_{i=1}^{m} E\{C_i[\hat{\theta}_i(y),\Theta_i]\,|\,Y = y\} \tag{4.2.37}$$

式(4.2.37)本质上是求解一个m阶估计问题，即选择$\hat{\theta}_i(y)$($\hat{\theta}(y)$的第i个元素)使$E\{C_i[\hat{\theta}_i(y),\ \Theta_i]\,|\,Y=y\}$最小。

一个有效的代价函数的例子就是误差的欧氏范数的平方，形如式(4.2.36)

$$C[\boldsymbol{a},\boldsymbol{\theta}] = \|\boldsymbol{a} - \boldsymbol{\theta}\|^2 = \sum_{i=1}^{m} (a_i - \theta_i)^2 \tag{4.2.38}$$

根据模型4.2.1的结论，与该代价函数对应的第i个元素的贝叶斯估计为$E\{\Theta_i\,|\,Y=y\}$；即贝叶斯估计为

$$\hat{\boldsymbol{\theta}}_{\mathrm{B}}(y) = E\{\boldsymbol{\Theta}\,|\,Y = y\} \tag{4.2.39}$$

即给定$Y=y$时$\boldsymbol{\Theta}$的条件均值。

满足式(4.2.36)的代价函数的另一个例子如下

⊖ 注意，在没有任何观测数据的情况下，最小均方误差估计值是μ，且最小均方误差为v^2，这与v^2d^2较小的估计结果近似对应。这再一次表明(正如下面的章节将会看到的)在没有任何先验信息情况下的最佳估计是$\hat{\theta}_1(\mathbf{y})$，其精确性为$1/d^2$。因此，在这两个极端情况下的估计结果是相当合理的。

$$C[\boldsymbol{a},\boldsymbol{\theta}] = \sum_{i=1}^{m} |a_i - \theta_i| \tag{4.2.40}$$

作为衡量 $\boldsymbol{a}$ 和 $\boldsymbol{\theta}$ 之间距离的量度，这个函数是除 $\|\boldsymbol{a}-\boldsymbol{\theta}\|$ 以外的另一种选择。结合模型4.2.2可以看出，上述代价函数在给定条件 $Y=y$ 下，得到的第 i 个元素的贝叶斯估计是 Θ_i 的条件中位数。

为了将最大后验估计的概念拓展到向量参数情况，思考式(4.2.36)给出的代价函数，其中 $C_i[a_i,\theta_i]$ 是式(4.2.14)的均匀代价函数。这导致了在给定条件 $Y=y$ 下，第 i 个元素是 Θ_i 的条件众数的估计值。然而，这种可分解的代价函数不是均匀代价函数对向量情形的最有意义的扩展。更有意义的形式是

$$C[\boldsymbol{a},\boldsymbol{\theta}] = \begin{cases} 1, & \max\limits_{1\leqslant i\leqslant m} |a_i - \theta_i| > \Delta \\ 0, & \max\limits_{1\leqslant i\leqslant m} |a_i - \theta_i| \leqslant \Delta \end{cases} \tag{4.2.41}$$

对于该代价函数，有

$$E\{C[\hat{\boldsymbol{\theta}}(\boldsymbol{y}),\boldsymbol{\Theta}]\,|\,Y=y\} = 1 - P(|\hat{\theta}_1(Y)-\Theta_1|)\leqslant\Delta,\cdots,|\hat{\theta}_m(Y)-\Theta_m|\leqslant\Delta\,|\,Y=y) \tag{4.2.42}$$

由式(4.2.42)，可以将给定条件 $Y=y$ 下的条件众数作为 $\boldsymbol{\Theta}$ 的近似最优估计，条件众数不同于一般向量，其第 i 个元素是在 $Y=y$ 条件下从代价中分解得到的条件众数 Θ_i。在 $Y=y$ 条件下选择 $\boldsymbol{\Theta}$ 的条件众数的估计是针对向量参数情形的 MAP 估计。注意，由 $\max\limits_{1\leqslant i\leqslant m}|a_i-\theta_i|\leqslant\Delta$ 确定的区域是一个以 $\boldsymbol{\theta}$ 为中心，棱长为 2Δ 的 m 维正方体。可以通过用其他形状(例如一个 m 维的球体，$\|\boldsymbol{a}-\boldsymbol{\theta}\|\leqslant\Delta$)取代这个正方体来定义类似的代价函数。然而，最大后验概率估计近似最优性仍然需要合适的平稳条件来保证。

在估计向量参数时，另一个更有用的代价函数是均方误差的一般形式。特别地，关注如下形式的代价函数

$$C[\boldsymbol{a},\boldsymbol{\theta}] = (\boldsymbol{a}-\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{a}-\boldsymbol{\theta}) \tag{4.2.43}$$

其中 $\boldsymbol{A}$ 是一个对称正定矩阵。注意，这个代价函数允许对不同参数的误差进行联合加权，这对某些应用而言是必需的，这是因为对于 $\boldsymbol{\Theta}$ 中的某一个参数，其估计的精度将影响到其他参数所需要的估计精度。

根据式(4.2.43)得到贝叶斯估计为

$$\begin{aligned} &E\{(\hat{\boldsymbol{\theta}}(y)-\boldsymbol{\Theta})^{\mathrm{T}}\boldsymbol{A}(\hat{\boldsymbol{\theta}}(y)-\boldsymbol{\Theta})\,|\,Y=y\} \\ &= [\hat{\boldsymbol{\theta}}(y)]^{\mathrm{T}}\boldsymbol{A}\,\hat{\boldsymbol{\theta}}(y) - 2[\hat{\boldsymbol{\theta}}(y)]^{\mathrm{T}}\boldsymbol{A}E\{\boldsymbol{\Theta}\,|\,Y=y\} + E\{\boldsymbol{\Theta}^{\mathrm{T}}\boldsymbol{A}\Theta\,|\,Y=y\} \end{aligned} \tag{4.2.44}$$

因为式(4.2.44)是关于 $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ 的二次函数，所以它在关于 $\hat{\boldsymbol{\theta}}(\boldsymbol{y})$ 的梯度趋于零时达到最小值，因此可直接得出

$$\nabla_{\hat{\boldsymbol{\theta}}(y)}E\{C[\hat{\boldsymbol{\theta}}(\boldsymbol{y}),\boldsymbol{\Theta}]\,|\,Y=y\} = 2\boldsymbol{A}\,\hat{\boldsymbol{\theta}}(y) - 2\boldsymbol{A}E\{\boldsymbol{\Theta}\,|\,Y=y\} \tag{4.2.45}$$

所以，根据式(4.2.43)，贝叶斯估计 $\hat{\boldsymbol{\theta}}_{\mathrm{B}}$ 满足

$$2\boldsymbol{A}\,\hat{\boldsymbol{\theta}}_{\mathrm{B}}(y) = 2\boldsymbol{A}E\{\boldsymbol{\Theta}\,|\,Y=y\} \tag{4.2.46}$$

将式(4.2.46)左乘 $\frac{1}{2}\boldsymbol{A}^{-1}$ 可以得到 $\hat{\boldsymbol{\theta}}_{\mathrm{B}}(y)=E\{\boldsymbol{\Theta}\,|\,Y=y\}$。

由式(4.2.43)可见，二次代价准则得到的条件均值向量等效为贝叶斯估计，与 A 的选

择无关。当然，贝叶斯风险的取值依赖于 $\boldsymbol{A}$，且可表示为(见习题 10)

$$r(\hat{\theta}_B) = \text{tr}\{\boldsymbol{A}E\{\text{cov}(\Theta|Y)\}\} \tag{4.2.47}$$

其中 tr{·}表示迹算子(对角线各项的总和)，cov($\boldsymbol{\Theta}|Y$)是 $Y=y$ 时 $\boldsymbol{\Theta}$ 的条件协方差矩阵。注意，均方误差范数是式(4.2.47)在 $\boldsymbol{A}=\boldsymbol{I}$ 时的一个特殊情况，所以 $r(\hat{\boldsymbol{\theta}}_B)$ 简化为 $E\{\text{cov}(\Theta|Y)\}$的迹。

例 4.2.3 联合高斯观测中的高斯向量估计

考虑 $\Gamma=\mathbf{R}^n$，$\Lambda=\mathbf{R}^m$ 的情况，Y 和 $\boldsymbol{\Theta}$ 分别是以 $\boldsymbol{\mu}_Y$ 和 $\boldsymbol{\mu}_\Theta$ 为均值向量、协方差矩阵为 $\boldsymbol{\Sigma}_Y$ 和 $\boldsymbol{\Sigma}_\Theta$、互协方差矩阵为 $\boldsymbol{\Sigma}_{Y\Theta} \triangleq E\{(\boldsymbol{Y}-\boldsymbol{\mu}_Y)(\boldsymbol{\Theta}-\boldsymbol{\mu}_\Theta)^{\mathrm{T}}\}$的联合高斯分布。即假设

$$\begin{bmatrix}\boldsymbol{Y}\\ \boldsymbol{\Theta}\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{\mu}_Y\\ \boldsymbol{\mu}_\Theta\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_Y & \boldsymbol{\Sigma}_{Y\Theta}\\ \boldsymbol{\Sigma}_{\Theta Y} & \boldsymbol{\Sigma}_\Theta\end{bmatrix}\right) \tag{4.2.48}$$

其中 $\boldsymbol{\Sigma}_{\Theta Y}=\boldsymbol{\Sigma}_{Y\Theta}^{\mathrm{T}}$。

在这个模型中，可直接得到给定 $\boldsymbol{Y}=\boldsymbol{y}$ 下 $\boldsymbol{\Theta}$ 的条件分布同样服从高斯分布，其条件平均数$\hat{\boldsymbol{\mu}}(y)$以及条件协方差矩阵$\hat{\boldsymbol{\Sigma}}$由下式给出

$$\begin{cases}\hat{\boldsymbol{\mu}}(\boldsymbol{y}) = \boldsymbol{\mu}_\Theta + \boldsymbol{\Sigma}_{\Theta Y}\boldsymbol{\Sigma}_Y^{-1}(\boldsymbol{y}-\boldsymbol{\mu}_Y)\\ \hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_\theta - \boldsymbol{\Sigma}_{\Theta Y}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\Sigma}_{Y\Theta}\end{cases} \tag{4.2.49}$$

从这条性质中，可以得到所有在模型 4.2.4 中讨论过的最优估计。特别地，其中的条件均值估计与式(4.2.49)中的$\hat{\boldsymbol{\mu}}(\boldsymbol{y})$相等。此外，由于多变量高斯分布的众数在其均值处取得，最大后验概率同样由$\hat{\boldsymbol{\mu}}(\boldsymbol{y})$给出。与此同时，因为给定条件 $\boldsymbol{Y}=\boldsymbol{y}$ 下 $\boldsymbol{\Theta}$ 服从高斯分布，意味着 Θ_i 在该条件下是边缘高斯分布，给定 $\boldsymbol{Y}=\boldsymbol{y}$ 时 Θ_i 的临界众数和中位数存在于$\hat{\mu}_i(\boldsymbol{y})$中，即$\hat{\boldsymbol{\mu}}(\boldsymbol{y})$的第 i 个元素。综上所述，根据在模型 4.2.4 中的讨论，$\hat{\boldsymbol{\mu}}(\boldsymbol{y})$给出了所有可能的最优估计。注意，这个估计与 $\boldsymbol{y}$ 成线性关系(更恰当地说是映射)，如果能够确定 $\boldsymbol{\Sigma}_Y^{-1}$，那么就可以很容易地计算出该最优估计。下文中将会对这个问题作进一步阐述。

通过式(4.2.47)得到式(4.2.43)的二次代价函数，最小贝叶斯风险同样可以容易地计算出来。特别地，由于 cov($\boldsymbol{\Theta}|\boldsymbol{Y}$)=$\hat{\boldsymbol{\Sigma}}$与 $\boldsymbol{Y}$ 无关，因此 $E\{\text{cov}(\boldsymbol{\Theta}|\boldsymbol{Y})\}=\hat{\boldsymbol{\Sigma}}$，并且最小贝叶斯风险变为

$$r(\hat{\boldsymbol{\theta}}_B) = \text{tr}\{A\,\hat{\boldsymbol{\Sigma}}\} = \text{tr}\{A\boldsymbol{\Sigma}_\theta\} - \text{tr}\{A\boldsymbol{\Sigma}_{\Theta Y}\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\Sigma}_{Y\Theta}\} \tag{4.2.50}$$

注意，$\hat{\boldsymbol{\Sigma}}=E\{(\boldsymbol{\Theta}-\hat{\boldsymbol{\theta}}_B(\boldsymbol{Y}))(\boldsymbol{\Theta}-\hat{\boldsymbol{\theta}}_B(\boldsymbol{Y}))^{\mathrm{T}}\}$，所以$\hat{\boldsymbol{\Sigma}}$是估计误差 $\boldsymbol{\Theta}-\hat{\boldsymbol{\theta}}_B(\boldsymbol{Y})$的协方差矩阵。

通用高斯模型的一个特殊情况就是线性观测模型

$$\boldsymbol{Y} = \boldsymbol{H\Theta} + \boldsymbol{N} \tag{4.2.51}$$

其中 $\boldsymbol{\Theta}\sim\mathcal{N}(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$，$\boldsymbol{N}\sim\mathcal{N}(0, \boldsymbol{\Sigma})$，$\boldsymbol{H}$ 是一个固定的 $n\times m$ 矩阵，$\boldsymbol{\Theta}$ 和 $\boldsymbol{N}$ 是相互独立的。这个模型出现在许多实际应用中。例如，在例 4.2.2 的模型中，希望估计信号的幅度就是这种形式，其中 $m=1$，$\boldsymbol{H}=\boldsymbol{s}$。此外，如果把 $\Theta_1, \cdots, \Theta_m$ 看做随机信号的样本，则有

$$Y_k = \sum_{j=1}^{m} h_{k,j}\Theta_j + N_k, \quad k = 1,\cdots,n \tag{4.2.52}$$

这就是一个由线性滤波后的信号加上加性噪声组成的观测序列，例如，观测量可由信号通过有限带宽或者其他线性失真的信道得到。在此情况下，对 $\boldsymbol{\Theta}$ 的估计就等价为通道均衡。该模型的一个更深入的应用将在第 5 章 Kalman-Bucy 滤波器的相关内容中讨论。

针对这个模型，显然在给定 $\boldsymbol{\mu}_\theta$ 和 $\boldsymbol{\Sigma}_\Theta$ 情况下，$\boldsymbol{Y}$ 和 $\boldsymbol{\Theta}$ 是联合高斯密度分布，$\boldsymbol{\mu}_Y=\boldsymbol{H}\boldsymbol{\mu}_\theta$，$\boldsymbol{\Sigma}_Y=\boldsymbol{H}\boldsymbol{\Sigma}_\theta\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{\Sigma}$，$\boldsymbol{\Sigma}_{\Theta Y}=\boldsymbol{\Sigma}_\theta\boldsymbol{H}^{\mathrm{T}}$。因此可以得到贝叶斯估计

$$\hat{\boldsymbol{\mu}}(\boldsymbol{y})=\boldsymbol{\mu}_\theta+\boldsymbol{\Sigma}_\Theta\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{\Sigma}_\theta\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{\Sigma})^{-1}(\boldsymbol{y}-\boldsymbol{H}\boldsymbol{\mu}_\theta) \tag{4.2.53}$$

其误差协方差矩阵为

$$\hat{\boldsymbol{\Sigma}}=\boldsymbol{\Sigma}_\theta-\boldsymbol{\Sigma}_\theta\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{\Sigma}_\theta\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{\Sigma})^{-1}\boldsymbol{H}\boldsymbol{\Sigma}_\theta \tag{4.2.54}$$

注意，式(4.2.53)的计算中包含了一个 $n\times n$ 矩阵的逆，其复杂度大约是 n^3(除非这个矩阵有某种特殊的结构)。在某些时候，可以通过简单的矩阵恒等运算来降低矩阵求逆的计算复杂度

$$\boldsymbol{\Sigma}_\theta H^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{\Sigma}_\theta\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{\Sigma})^{-1}=(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{H}+\boldsymbol{\Sigma}_\theta^{-1})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1} \tag{4.2.55}$$

如果 $\boldsymbol{\Sigma}^{-1}$ 是已知的(如 $\boldsymbol{\Sigma}=\sigma^2 I$)且 $m<n$，则式(4.2.55)中右边的矩阵相比于左边的矩阵要更容易计算。

在第 5 章中，式(4.2.53)和式(4.2.54)将会被用于推导 Kalman-Bucy 滤波器。根据线性观测的统一模型，可重做例 4.2.2。在这种情况下，有 $m=1$，$\boldsymbol{H}=\boldsymbol{s}$，$\boldsymbol{\mu}_\theta=\mu$，$\boldsymbol{\Sigma}_\theta=v^2$。将这些量代入式(4.2.53)和式(4.2.54)，并且运用式(4.2.55)，可得

$$\hat{\boldsymbol{\mu}}(\boldsymbol{y})=\mu+(\boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{s}+1/v^2)^{-1}\boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{s}\mu)=\frac{v^2d^2\ \hat{\theta}_1(\boldsymbol{y})+\mu}{v^2d^2+1}$$

以及

$$r(\hat{\boldsymbol{\mu}})=\hat{\boldsymbol{\Sigma}}=v^2-(\boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{s}+v^{-2})^{-1}\boldsymbol{s}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{s}v^2=\frac{v^2}{v^2d^2+1}$$

与式(4.2.53)和式(4.2.53)所得结果相同。

4.3　非随机参数估计：基本理论结构

在 4.2 节中，考虑的问题是估计某个随机参数，并用它来表示观测空间的一类分布。与之对应的问题是：待估计的参数(用来表示观测统计量族)不是随机变量，但它却是未知的。特别地，获得的有关参数的先验信息不足以将其划归为某种先验概率分布，但仍希望对此类参数估计问题进行有效处理。

假设观测为 $Y\in\Gamma$，同时 Y 的分布由参数 θ 所描述的分布族(Γ，G)中的某个分布确定。同前文一样，用$\{P_\theta;\ \theta\in\Lambda\}$来表示分布函数的集合，且假设参数 θ 是实数。对于 θ 的实际值，仅知道它在 Λ 中取值。简而言之，所要解决的问题是：给定一个观测 $Y=y$，θ 的最优估计是什么？参考 4.2 节中的证明过程，答案可能是寻找一个估计$\hat{\theta}(y)$使得某种平均性能评价指标最小。本章后续内容将只考虑将平方误差作为代价函数，与之相关的结论可直接推广到其他形式的代价函数。在缺乏 Λ 的先验信息情况下，平均代价只能通过计算给定 θ 下 Y 的分布来确定，即只能使用条件风险函数 $R_\theta(\hat{\theta})\triangleq E_\theta\{(\hat{\theta}(Y)-\theta)^2\}$，$\theta\in\Lambda$。

与第 2 章的情况一致，不能期望对于所有 $\theta\in\Lambda$ 都让 $R_\theta(\theta)$最小化。用一个简单的例子来说明这一问题：如果将平方误差作为代价，对于任何 θ(比如 θ_0)，如果对于所有的 $y\in\Gamma$ 都选择 θ_0 作为估计子$\hat{\theta}(y)$，条件均方误差可以为零；但是当 θ_0 不在 θ 的实际值附近时，前述估计的性能将会很差。显然，条件均方误差并非衡量非随机参数估计子性能的合适标准，除非限制估计器是合理的[比如，将估计$\hat{\theta}(y)\equiv\theta_0$ 排除在外]。

对于 θ 估计值的一个合理约束条件是它的期望等于实际参数，即

$$E_\theta\{\hat{\theta}(Y)\} = \theta, \quad \theta \in \Lambda \tag{4.3.1}$$

上述估计子称为无偏估计。在这个限制之下，条件均方误差变为估计子在 P_θ 下的方差，并且对于每一个 $\theta \in \Lambda$，如果无偏估计能最小化均方误差，则称之为最小方差无偏估计子(MVUE)。

本节通过讨论非随机参数估计问题来描述 MVUE。首先，需要从充分性的概念开始，它的定义如下(除非另行提及，假设 Λ 是一般性的集合，即不要求一定是 $\mathbf{R}$ 的子集)。

定义 4.3.1　充分性

假设 Δ 是任意集合，D 是关于 Δ 的事件类。对于 $Y \sim P_\theta$ 和 $\theta \in \Lambda$，如果 $T(Y)$ 的分布与 θ 无关，则称 T：$(\Gamma, \mathcal{G}) \to (\Delta, D)$ 是 $\{P_\theta;\ \theta \in \Lambda\}$ 的充分统计量。(如果 $\{P_\theta;\ \theta \in \Lambda\}$ 已经确定，则可简单地称 T 关于 θ 是充分的。)

注意，θ 仅仅通过分布 P_θ 来影响观测值，所以只能通过观测 Y 的统计特性来估计 θ。一旦 $T(Y)$ 能够去除观测 Y 对于 θ 的依赖，就能够推断出 $T(Y)$ 包含了 Y 中所有用于估计 θ 的有用信息——这也是“充分性”的本质。

需要注意的是：任何同观测值一一对应的映射都是 θ 的平凡充分统计量，即对于任意给定的估计模型都存在多个充分统计量。然而，希望找到一个能够尽可能减少观测样本的充分统计量。为此，需要定义最小充分性，具体如下。

定义 4.3.2　最小充分性

如果函数 T 是任何其他关于 $\{P_\theta;\ \theta \in \Lambda\}$ 的充分统计量的函数，那么称在 $(\Gamma, \mathcal{G})$ 上的函数 T 对于 $\{P_\theta;\ \theta \in \Lambda\}$ 是最小充分的。

换句话说，最小充分统计量在不损失关于 θ 的信息的情况下最大限度地减小了观测量。但是，很多估计问题不存在最小充分统计量，并且即使存在最小统计量，也很难求解。

不过，基于下面的结论，通常可以很容易地找到可用的(尽管不一定是最小的)充分统计量。

命题 4.3.1　因子分解定理

假设与 $\{P_\theta;\ \theta \in \Lambda\}$ 相对应的密度族为 $\{p_\theta;\ \theta \in \Lambda\}$，统计量 T 是 θ 的充分统计量的充要条件是存在函数 g_θ 和 h 使得下式对于所有 $y \in \Gamma$ 和 $\theta \in \Lambda$ 成立

$$p_\theta(y) = g_\theta[T(y)]h(y) \tag{4.3.2}$$

证明　以下仅给出 Γ 是离散情况下的证明。这是因为离散情况下的证明过程能充分阐明本命题的基本思想，并且不需要引入证明一般性结果时所需要的技巧。对于一般情况下的证明过程，读者可以在 Lehmann(1986)的书中查阅到相关内容。

假设 Γ 是离散的，并且 $\{p_\theta;\ \theta \in \Lambda\}$ 和函数 T 满足式(4.3.2)。对于 $Y \sim P_\theta$，如果给定 $T(Y)=t$，用 $p_\theta(y|t)$ 来表示给定 $T(Y)=t$ 和 $Y \sim P_\theta$ 时的概率密度，则根据贝叶斯公式有

$$p_\theta(y|t) \triangleq P_\theta\{Y = y|T(Y) = t\} = \frac{P_\theta\{T(Y) = t|Y = y\}P_\theta(Y = y)}{P_\theta(T(Y) = t)} \tag{4.3.3}$$

因为 $T(Y)=t$ 和 $T(Y) \neq t$ 时，$P_\theta\{T(Y)=t|Y=y\}$ 分别等于 1 和 0，并且由于 $P_\theta(Y=y)=p_\theta(y)$，所以式(4.3.3)可化简为

$$p_\theta(y|t)=\begin{cases}p_\theta(y)/P_\theta(T(Y)=t), & T(y)=t\\ 0, & T(y)\neq t\end{cases} \tag{4.3.4}$$

利用 $P_\theta(T(Y)=t)=\sum_{\{y|T(Y)=t\}}p_\theta(y)$ 并结合式(4.3.2)后，可以得到

$$P_\theta(T(y)=t)=\sum_{\{y|T(Y)=t\}}g_\theta[T(y)]h(y)=g_\theta(t)\sum_{\{y|T(y)=t\}}h(y)$$

和 $p_\theta(y)=g_\theta[T(y)]h(y)=g_\theta(t)h(y)$。进而可将式(4.3.4)写成

$$p_\theta(y|t)=\begin{cases}h(y)/\sum_{y|T(y)=t\}}h(y), & T(y)=t\\ 0, & T(y)\neq t\end{cases}$$

显然，上式不依赖于 θ，即 T 是 $\{P_\theta;\ \theta\in\Lambda\}$ 的充分统计量。换言之，如果式(4.3.2)成立，则 T 是充分统计量。

为了证明式(4.3.2)是 T 为充分统计量的必要条件，假设 T 是 θ 的任意充分统计量。由式(4.3.4)可以得到

$$p_\theta(y)=p_\theta[y|T(y)]P_\theta[T(Y)=T(y)] \tag{4.3.5}$$

一方面，T 是 θ 的充分统计量，$p_\theta[y|T(y)]$ 仅依赖于 y；另一方面，$P_\theta[T(Y)=T(y)]$ 仅是 $T(y)$ 和 θ 的函数。在定义 $h(y)\triangleq p_\theta[y|T(y)]$ 和 $g_\theta[T(y)]\triangleq P_\theta[T(Y)=T(y)]$ 后，可以看出，式(4.3.5)就是(4.3.2)中的因子分解形式。至此，完成了 Γ 是离散情况下的证明过程。 □

为了阐明命题 4.3.1，考虑下面简单的例子。

例 4.3.1　假设检验的充分统计量

考虑假设检验问题 $\Lambda=\{0,\ 1\}$，其概率密度分别为 p_0 和 p_1，注意

$$p_\theta(y)=\begin{cases}p_0(y), & \theta=0\\ \dfrac{p_1(y)}{p_0(y)}p_0(y), & \theta=1\end{cases}$$

可见，$p_\theta(y)=g_\theta[T(y)]h(y)$，其中，$h(y)=p_0(y)$，$T(y)=p_1(y)/p_0(y)\triangleq L(y)$，且

$$g_\theta(t)=\begin{cases}1, & \theta=0\\ t, & \theta=1\end{cases}$$

根据因子分解定理易知：似然比 $L(y)$ 是二项假设检验问题的充分统计量。似然比是非常实用的充分统计量，因为它不但具有低维(只有一维)特性，而且与 Γ 的性质无关。显然，在第 2 章中所有有关 $\Lambda=\{0,\ 1\}$ 的仅依赖于观测数据 y 的最优检验均以最大似然比 $L(y)$ 作为充分统计量。 ◀

下面的 Rao-Blackwell 定理描述了充分统计量在求解实参数的无偏估计子时的作用。这里，允许 Λ 是任意的，但是假设以 θ 为变量的函数 g 是实函数。

命题 4.3.2　Rao-Blackwell 定理

假设 $\hat{g}(y)$ 是 $g(\theta)$ 的无偏估计，并且 T 是 θ 的充分统计量。定义 $\tilde{g}[T(y)]$ 为

$$\tilde{g}[T(y)]=E_\theta\{\hat{g}(Y)|T(Y)=T(y)\}$$

那么 $\tilde{g}[T(y)]$ 也是 $g(\theta)$ 的无偏估计。进一步有

$$\mathrm{var}_\theta(\tilde{g}[T(Y)])\leqslant\mathrm{var}_\theta(\hat{g}(Y))$$

当且仅当 $P_\theta(\hat{g}(y)=\widetilde{g}[T(Y)])=1$ 时等号成立。

证明 首先，利用 T 是充分统计量的性质（即 $T(y)$ 的分布不依赖于 θ），$\widetilde{g}$ 的期望不依赖于参数 θ。为了说明 $\widetilde{g}$ 是无偏的，利用

$$E_\theta\{\widetilde{g}[T(Y)]\} = E_\theta\{E_\theta\{\hat{g}(Y)\mid T(y)\}\} = E_\theta\{\hat{g}(Y)\} = g(\theta)$$

其中使用了全期望公式 $E\{E\{X|Z\}\}=E\{X\}$ 来得到第二个等式，并利用 $\hat{g}$ 的无偏性得到第三个等式。

为了证明 $\mathrm{var}_\theta(\widetilde{g}[T(Y)])\leqslant\mathrm{var}_\theta(\hat{g}(Y))$，首先有

$$\mathrm{var}_\theta(\widetilde{g}[T(Y)]) = E_\theta\{[\widetilde{g}[T(Y)]]^2\} - g^2(\theta)$$

且

$$\mathrm{var}_\theta(\hat{g}(Y)) = E_\theta\{[\hat{g}(Y)]^2\} - g^2(\theta)$$

因此仅需要证明 $E_\theta\{[\widetilde{g}[T(Y)]]^2\}\leqslant E_\theta\{[\hat{g}(Y)]^2\}$，可得

$$\begin{aligned} E_\theta\{[\widetilde{g}[T(y)]]^2\} &= E_\theta\{[E_\theta\{\hat{g}(Y)\mid T(Y)\}]^2\} \\ &\leqslant E_\theta\{E_\theta\{[\hat{g}(Y)]^2\mid T(Y)\}\} = E_\theta\{[\hat{g}(Y)]^2\} \end{aligned} \tag{4.3.6}$$

其中，利用 Jensen 不等式可得 $[E_\theta\{\hat{g}(Y)\mid T(Y)\}]^2\leqslant E_\theta\{[\hat{g}(Y)]^2\mid T(Y)\}$，最后一个等号由全期望公式得到。Jensen 不等式取等号的条件是当且仅当 $P_\theta[\hat{g}(Y)=E_\theta\{\hat{g}(Y)\mid T(Y)\}\mid T(Y)]=1$。又因为 $\widetilde{g}[T(y)]\triangleq E_\theta\{\hat{g}(Y)\mid T(Y)\}$，即 Jensen 不等式取等号的条件与 $P_\theta[\hat{g}(Y)=\widetilde{g}[T(Y)]]=1$ 等价。至此，完成了命题 4.3.2 的证明。 □

从 Rao-Blackwell 定理可以得出：利用充分统计量 T，可以通过对其施加条件 $T(Y)$ 来改善任何无偏估计子，只要这些估计子不是 T 的函数。进一步地，根据 Rao-Blackwell 定理，如果 T 对于 θ 是充分的，并且 $g(\theta)$ 以 T 为变量的无偏估计子仅有一个，那么该估计子就是 $g(\theta)$ 的 MVUE。为了证明这一点，假设 $g^*[T(y)]$ 是 $T(y)$ 的唯一满足条件 $E_\theta\{g^*[T(y)]\}=g(\theta)$ 的函数。令 $\hat{g}(y)$ 是 $\hat{g}(\theta)$ 的任意无偏估计子。那么，基于 Rao-Blackwell 定理，$\widetilde{g}[T(y)]\triangleq E_\theta\{\hat{g}(Y)\mid T(Y)=T(y)\}$ 是 $g(\theta)$ 的无偏估计子，并且它是 $T(y)$ 的函数。因此，基于 g^* 的唯一性，一定有 $g^*=\widetilde{g}$。Rao-Blackwell 定理同时也阐述了 $\mathrm{var}_\theta(\widetilde{g}[T(Y)])\leqslant\mathrm{var}_\theta(\hat{g}(Y))$。由于 $\hat{g}$ 是任意的，得出 $\mathrm{var}_\theta(g^*[T(Y)])\leqslant\mathrm{var}_\theta(\hat{g}(Y))$ 对于 $g(\theta)$ 的任何无偏估计子都成立，即 $g^*[T(y)]$ 是 $g(\theta)$ 的 MVUE。

综合上述分析，如果能够找到某个具有唯一无偏估计子 $g^*[T(y)]$ 的充分统计量 T，那么就可以通过它来构造 $g(\theta)$ 的 MVUE。为了找到这样的统计量，需要引进完备性的概念。

定义 4.3.3 完备性

族 $\{P_\theta;\ \theta\in\Lambda\}$ 具有完备性的条件是，$E_\theta\{f(Y)\}=0$ 对于所有 $\theta\in\Lambda$ 成立，等价于 $P_\theta[f(Y)=0]=1$ 对于所有 $\theta\in\Lambda$ 成立。

这种完备性的概念同 $\mathbf{R}^n$ 中向量集合的完备性的概念十分相似。为了说明这点，考虑 Γ 是有限集合 $\{\gamma_1,\ \cdots,\ \gamma_n\}$ 的情况。在此种情况下，Γ 上的任意函数 f 可以写成

$$E_\theta\{f(Y)\} = \boldsymbol{f}^{\mathrm{T}}\boldsymbol{p}_\theta$$

其中 $\boldsymbol{f}=[f(\gamma_1),\ f(\gamma_2),\ \cdots,\ f(\gamma_n)]^{\mathrm{T}}$，并且 $\boldsymbol{p}_\theta=[p_\theta(\gamma_1),\ p_\theta(\gamma_2),\ \cdots,\ p_\theta(\gamma_n)]^{\mathrm{T}}$。假设 $p_\theta(\gamma_i)>0$ 对于所有 $\theta\in\Lambda$ 和 $i=1,\ \cdots,\ n$ 都成立，$\{P_\theta;\ \theta\in\Lambda\}$ 的完备性定义为 $\boldsymbol{f}^{\mathrm{T}}\boldsymbol{p}_\theta=0$ 对

于所有 $\theta \in \Lambda$ 成立，这意味着 $\boldsymbol{f}$ 是一个全零的 n 维向量。即如果 $\mathbf{0}$ 是唯一的垂直于 $\{\boldsymbol{p}_\theta;\ \theta \in \Lambda\}$ 中所有元素的向量，则 $\{P_\theta;\ \theta \in \Lambda\}$ 是完备的。显然，这是 $\mathbf{R}^n$ 上向量集合 $\{p_\theta;\ \theta \in \Lambda\}$ 具有完备性的一般性定义（向量空间 $\mathbf{R}^n$ 的完备集合张成空间 $\mathbf{R}^n$）。类似的结论对于更加一般的观测空间也成立。

为了说明完备性的概念，考虑例 4.3.2。

例 4.3.2　二项分布的完备性

假设 $\Gamma=\{0,\ 1,\ \cdots,\ n\}$，$\Lambda=(0,\ 1)$，并且

$$p_\theta(y) = \frac{n!}{y!(n-y)!}\theta^y(1-\theta)^{n-y}, \quad y = 0,\cdots,n, \quad 0 < \theta < 1$$

对于 Γ 上的任意函数 f 有

$$E_\theta\{f(Y)\} = \sum_{y=0}^{n} \frac{n!}{y!(n-y)!} f(y)\theta^y(1-\theta)^{n-y} = (1-\theta)^n \sum_{y=1}^{n} a_y x^y$$

其中

$$a_y \triangleq \frac{n!}{y!(n-y)!} f(y), \quad y = 0,\cdots,n$$

且

$$x \triangleq \theta/(1-\theta)$$

对于所有 $\theta \in \Lambda$，$E_\theta\{f(Y)\}=0$ 成立等价于

$$\sum_{y=0}^{n} a_y x^y = 0, \quad x > 0 \tag{4.3.7}$$

因为函数 $\sum_{y=0}^{n} a_y x^y$ 是 n 阶多项式，且除非所有系数都为零，否则只有当 $f(y)=0$，$y=0$，$\cdots$，n 时式(4.3.7)才成立，进而 $\{p_\theta;\ \theta \in \Lambda\}$ 是完备的。事实上，对于任意的拥有至少 $n+1$ 个非零参数的 Λ，上述完备性都成立。◀

完备性的概念和充分性的概念联系紧密。为了说明这点，假设 T 对整个分布族 $\{P_\theta;\ \theta \in \Lambda\}$ 是充分的，同时为了简便，假设 $E_\theta\{|Y|\}<\infty$ 对于所有 $\theta \in \Lambda$ 成立。定义函数 $f(y)$ 为

$$f(y) = y - E_\theta\{Y \mid T(Y) = T(y)\}$$

因为 T 是充分的，函数 f 并不依赖于 θ，所以对于所有 $\theta \in \Lambda$ 有下面的表达式成立

$$E_\theta\{f(Y)\} = E_\theta\{Y\} - E_\theta\{E_\theta\{Y \mid T(Y)\}\} = E_\theta\{Y\} - E_\theta\{Y\} = 0$$

$\{P_\theta;\ \theta \in \Lambda\}$ 的完备性就隐含了 $P_\theta[Y=E_\theta\{Y \mid T(Y)\}]=1$ 对于所有 $\theta \in \Lambda$ 成立，实际上，$y=E_\theta\{Y \mid T(Y)=T(y)\}$。既然 $E_\theta\{Y \mid T(Y)=T(y)\}$ 是 $T(y)$ 的函数，后一个条件表明了 y 本身就是 $T(y)$ 的函数。又因为 $T(y)$ 显然是 y 的函数，所以可以推断出 $T(y)$ 一定是关于 y 的一一映射的函数，即 T 是一个平凡的充分统计量。综上可得：*如果 $\{P_\theta;\ \theta \in \Lambda\}$ 是完备的，那么不存在 θ 的非平凡充分统计量*，即观测量 Y 不能在不损失 θ 信息的同时降低数据量。

完备性在刻画 MVUE 的相关特性方面是一个非常有用的概念。为了阐明这一点，假设 T 对于 θ 是充分的，并且用 Q_θ 表示当 $Y \sim P_\theta$ 时 $T(Y)$ 的分布。如果 $\{Q_\theta;\ \theta \in \Lambda\}$ 是完备

的，那么 T 就被称为完备充分统计量[㊀]。假设 T 是完备的，记$\widetilde{g}[T(y)]$和 $g^*[T(y)]$为 $T(y)$的任意函数且是 $g(\theta)$的无偏估计子，有

$$E_\theta\{\widetilde{g}[T(Y)]-g^*[T(Y)]\}=E_\theta\{\widetilde{g}[T(Y)]\}-E_\theta\{g^*[T(Y)]\}=g(\theta)-g(\theta)=0$$

上式对于所有 $\theta\in\Lambda$ 成立。因此，基于 T 的完备性，有 $P_\theta(\widetilde{g}[T(Y)]=g^*[T(Y)])=1$ 对于所有 $\theta\in\Lambda$ 成立，即$\widetilde{g}[T(Y)]$和 $g^*[T(Y)]$是同一个估计子。既然$\widetilde{g}$和 g^* 是任意选取的，那么可以得出结论：*以完备充分统计量作为函数的任何无偏估计子是唯一的，并且它就是* MVUE。至此，可以给出求解 MVUE 的步骤：

1）求解$\{P_\theta;\ \theta\in\Lambda\}$的完备充分统计量 T；

2）求解 $g(\theta)$的任意无偏估计子$\hat{g}(y)$；

3）$\widetilde{g}[T(y)]\triangleq E_\theta\{\hat{g}(Y)\mid T(Y)=T(y)\}$就是 $g(\theta)$的 MVUE。

上述计算步骤中，第一步看起来似乎最困难，因为第二步通常十分简单，第三步可以通过概率积分直接求得。然而，在很多实际估计模型中，第一步却十分简单。为了进一步说明这一点，首先给出指数族的定义。

定义 4.3.4　指数族

一类分布$\{P_\theta;\ \theta\in\Lambda\}$被称为指数族的条件是，如果存在实值函数 C，Q_1，…，Q_m，T_1，…，T_m 和 h，使得 P_θ 的密度表达式对所有 $\theta\in\Lambda$ 和 $y\in\Gamma$ 都成立

$$p_\theta(y)=C(\theta)\exp\Big\{\sum_{l=1}^{m}Q_l(\theta)T_l(y)\Big\}h(y)\tag{4.3.8}$$

很多实际分布都能写成指数族的形式，其中包括高斯分布、泊松分布、拉普拉斯分布、二项分布、几何分布和这些分布的某些多分量形式。指数族在最小方差无偏估计中占有重要地位，这要归功于下面将要证明的定理。

命题 4.3.3　指数族完备定理

假设 $\Gamma=\mathbf{R}^n$，$\Lambda\subset\mathbf{R}^m$，并且每一个 P_θ 都有密度函数 p_θ 为

$$p_\theta(y)=C(\theta)\exp\Big\{\sum_{l=1}^{m}\theta_l T_l(y)\Big\}h(y)\tag{4.3.9}$$

其中 C、Q_1，…，Q_m、T_1，…，T_m 和 h 都是实值函数[㊁]*，那么 $T(y)=[T_1(y),\ \cdots,\ T_m(y)]$ 是$\{P_\theta;\ \theta\in\Lambda\}$的完备充分统计量的条件是 Λ 包括 m 维矩形区域。*

证明　对于上述命题的完整证明可以参见 Lehman(1986)。具体证明步骤可概括如下。

首先，由因子分解定理可知 T 对于 θ 是充分的(命题 4.3.1)，因此仅需要证明 T 的完备性。Y 服从式(4.3.9)的分布形式，直接得到 $T(Y)$的密度(在 $\mathbf{R}^m$ 上)为

$$g_\theta(t)=C(\theta)\exp\Big\{\sum_{l=1}^{m}\theta_l t_l\Big\}h_T(t)\tag{4.3.10}$$

其中 h_T 是 t 的实值函数。假设 f 是 $\mathbf{R}^m$ 上的实值函数且满足 $E_\theta\{f[T(Y)]\}=0$，有

$$E_\theta\{f[T(Y)]\}=C(\theta)\int_{\mathbf{R}^m}f(t)\exp\Big\{\sum_{l=1}^{m}\theta_l t_l\Big\}h_T(t)\mu(\mathrm{d}t)\tag{4.3.11}$$

㊀ 回顾上面的讨论，能够发现，在不损失 θ 信息的情况下，对于观测维度的降低不能超过 $T(Y)$。实际上，T 必须是$\{P_\theta;\ \theta\in\Lambda\}$的极小充分统计量。这是因为，对于任何充分统计量 T'，由于完备性，一定有 $T=E_\theta\{T\mid T'\}$。因此，T 是 T'的函数。

㊁ 注意通过改变参数，式(4.3.8)可以写为式(4.3.9)的形式。

假设 Λ 包括了 m 维矩形区域 $J=\{\theta|a_1\leqslant\theta_1\leqslant b_1,\ a_2\leqslant\theta_2\leqslant b_2,\ \cdots,\ a_m\leqslant\theta_m\leqslant b_m\}$。仅通过参数的简单变换总是能使矩形区域变为 $J'=\{\theta|-1\leqslant\theta_1\leqslant1,\ -1\leqslant\theta_2\leqslant1,\ \cdots,\ -1\leqslant\theta_m\leqslant1\}$。将式(4.3.11)中的 θ_l 替换为($\theta_l+\mathrm{i}u_l$，$l=1$，…，m)后，原式转化为复数变量函数。可以证明该函数在区域 $C=\{\theta+\mathrm{i}u_l\,|-1\leqslant\theta_l\leqslant1,\ -\infty\leqslant u_l\leqslant\infty,\ l=1,\ \cdots,\ m\}$上解析，那么对于 J'中所有实元素都等于零的情况，表明它在 C 确定的条带中取值都为零。特别地，在区域 $C'=\{\theta+\mathrm{i}u_l\,|\theta_l=0,\ -\infty\leqslant u_l\leqslant\infty,\ l=1,\ \cdots,\ m\}$中此函数取零，也就是下式对于所有 $u\in\mathbf{R}^m$ 都成立。

$$C(\theta)\int_{\mathbf{R}^m}f(t)\exp\left\{\mathrm{i}\sum_{l=1}^{m}u_lt_l\right\}h_T(t)\mu(\mathrm{d}t)=0\tag{4.3.12}$$

注意，函数(4.3.12)的左边是多维傅里叶变换，对所有 $\theta\in\Lambda$ 一致为零，表明被变换的函数对于所有 $\theta\in\Lambda$ 为零，或者等价地说 $P_\theta\{f(Y)=0\}=1$ 对于所有 $\theta\in\Lambda$ 成立，这表明 T 是完备的。至此，原命题的完备性证明完毕。 □

为进一步阐明命题 4.3.3，考虑如下例子。

例 4.3.3　信号幅度的最小方差无偏估计子

考虑如下模型

$$Y_k=N_k+\mu s_k,\quad k=1,\cdots,n$$

其中 N_1，…，N_n 是服从$\mathcal{N}(0,\ \sigma^2)$的独立同分布噪声样本，$\mathbf{s}=(s_1,\ \cdots,\ s_n)^{\mathrm{T}}$ 是已知信号，幅度参数为 μ。现在假设 σ^2 已知并且希望估计幅度参数 μ。$\mathbf{Y}$ 的概率密度如下式

$$\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(y_k-\mu s_k)^2\right\}=C(\theta_1)\exp\{\theta_1T_1(\mathbf{y})\}h(\mathbf{y})\tag{4.3.13}$$

其中定义

$$\theta_1=\mu/\sigma^2,T_1(y)=\sum_{k=1}^{n}s_ky_k,\quad C(\theta_1)=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{\theta_1^2\sigma^2}{2}\sum_{k=1}^{n}s_k^2\right\}$$

和

$$h(\mathbf{y})=\exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}y_k^2\right\}$$

假设 μ 是任意的实数，参数集合是 $\Lambda=\{\theta_1\,|-\infty<\theta_1<\infty\}=\mathbf{R}$。一维矩形区域就是一个闭区间，并且 Λ 显然包括这个区间，所以由命题 4.3.3 并结合式(4.3.13)，能得出 $T_1(y)$是 θ_1 的完备充分统计量。

此外，还希望能够估计 $\mu=g(\theta)=\sigma^2\theta_1$。因为 $E_\theta\{Y_1\}=\mu s_1$，在假设 $s_1\neq0$ 后，估计 $\hat{g}(y)=y_1/s_1$是 $g(\theta)$的无偏估计子。又因为 T_1 是完备的，所以估计是 MVUE，有

$$\tilde{g}[T_1(\mathbf{y})]=E_\theta\{\hat{g}(\mathbf{Y})\,|\,T_1(\mathbf{Y})=T_1(\mathbf{y})\}\tag{4.3.14}$$

为了计算式(4.3.14)，因为$\hat{g}(\mathbf{Y})$和 $T_1(\mathbf{Y})$都是 $\mathbf{Y}$ 的线性函数，且 $\mathbf{Y}$ 服从高斯分布，所以$\hat{g}(\mathbf{Y})$和 $T_1(\mathbf{Y})$服从联合高斯分布。进而有

$$E_\theta\{\hat{g}(\mathbf{Y})\}=\mu,E_\theta\{T_1(\mathbf{Y})\}=n\mu\,\overline{s^2}$$

$$\mathrm{var}_\theta\{\hat{g}(\mathbf{Y})\}=\sigma^2/s_1^2,\mathrm{var}_\theta\{T_1(\mathbf{Y})\}=n\sigma^2\,\overline{s^2}$$

和

$$\mathrm{cov}_\theta\{\hat{g}(\mathbf{Y}),T_1(\mathbf{Y})\}=\sigma^2$$

其中定义 $\overline{s^2} \triangleq (1/n)\sum_{k=1}^{n} s_k^2$。应用4.2节中的结论，可以将式(4.3.14)的条件均值写成如下形式

$$\begin{aligned}\widetilde{g}[T_1(\mathbf{y})] &= E_\theta\{\widetilde{g}(\mathbf{Y})\} + \mathrm{cov}_\theta\{\widetilde{g}(\mathbf{Y}), T_1(\mathbf{Y})\} \times [\mathrm{var}_\theta[T_1(\mathbf{Y})]]^{-1}[T_1(\mathbf{y}) - E_\theta\{T_1(\mathbf{Y})\}] \\ &= \mu + \sigma^2(n\sigma^2\,\overline{s^2})^{-1}[T_1(\mathbf{y}) - n\mu\,\overline{s^2}] = T_1(\mathbf{y})/n\,\overline{s^2} = \left(\sum_{k=1}^{n} s_k y_k\right)/n\,\overline{s^2} \end{aligned} \tag{4.3.15}$$

至此，已经建立了信号幅度 μ 的 MVUE。该估计子的方差是

$$\mathrm{var}_\theta(\widetilde{g}[T_1(\mathbf{Y})]) = \sigma^2/n\,\overline{s^2} \tag{4.3.16}$$

现在假设 μ 和 σ^2 都未知，μ 的取值范围是 $\mathbf{R}$，σ^2 的取值范围是(0，∞)，并且希望估计这两个参数。从式(4.3.16)中可以看出 σ^2 可用来衡量幅度估计的精确度。注意，在式(4.3.13)中定义的 $h(\mathbf{y})$ 是 σ^2 的函数，所以如果 σ^2 未知，则式(4.3.13)不满足指数族分布。然而，该分布函数可改写为

$$\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(y_k - \mu s_k)^2\right\} = C(\theta_1)\exp\{\theta_1 T_1(\mathbf{y}) + \theta_2 T_2(\mathbf{y})\}h(\mathbf{y}) \tag{4.3.17}$$

其中 θ_1 和 T_1 同式(4.3.13)中一样。如果定义

$$\theta = (\theta_1, \theta_2), \quad \theta_2 = -\frac{1}{2\sigma^2}, \quad T_2(\mathbf{y}) = \sum_{k=1}^{n} y_k^2$$

和

$$C(\theta) = \left(-\frac{\theta_2}{\pi}\right)^{n/2}\exp\left\{\frac{\theta_1^2}{4\theta_2}\sum_{k=1}^{n} s_k^2\right\}, \quad h(\mathbf{y}) \equiv 1$$

则值域$\{(\mu,\ \sigma^2)\,|\,\mu\in\mathbf{R},\ \sigma^2>0\}$对应于$\Lambda=\{(\theta_1,\ \theta_2)\,|\,\theta_1\in\mathbf{R},\ \theta_2<0\}$，显然包含矩形区域。因此，$T_1=(T_1,\ T_2)$是 θ 的完备充分统计量。

希望估计 $\mu = g_1(t) \triangleq -\theta_1/2\theta_2$ 和 $\sigma^2 = g_2(\theta) \triangleq -1/2\theta_2$。因为式(4.3.15)中给出的估计子在计算过程中不需要任何有关 σ^2 的先验知识，它是无偏的，且是 $T_1(\mathbf{y})$的函数(因此也是 $T(y)$的函数)。故即使在 σ^2 未知情况下，它也是 μ 的 MVUE。

为了求得 σ^2 的 MVUE，可以先寻找 σ^2 的无偏估计子，然后对它施加条件 $T(\mathbf{y})$。然而，直接求解一个以 T 为变量的无偏函数作为无偏估计子会更加简单。特别地，由于 $T_1(\mathbf{Y}) \sim \mathcal{N}(n\mu\,\overline{s^2},\ n\sigma^2\,\overline{s^2})$，有

$$E_\theta\{T_1^2(\mathbf{Y})\} = \mathrm{var}_\theta[T_1(\mathbf{Y})] + (E_\theta\{T_1(\mathbf{Y})\})^2 = n\sigma^2\,\overline{s^2} + n^2\mu^2(\overline{s^2})^2$$

类似地，有

$$E_\theta\{T_2(\mathbf{Y})\} = \sum_{k=1}^{n} E_\theta\{Y_k^2\} = \sum_{k=1}^{n}(\sigma^2 + \mu^2 s_k^2) = n\sigma^2 + n^2\mu^2\,\overline{s^2}$$

从这两个结果中得出$[T_2(\mathbf{Y}) - T_1^2(\mathbf{Y})/n\,\overline{s^2}]$的均值为

$$E_\theta\{T_2(\mathbf{Y})\} - E_\theta\{T_1^2(\mathbf{Y})/n\,\overline{s^2}\} = (n-1)\sigma^2 \tag{4.3.18}$$

因此函数$\widetilde{g}_2[T(y)] = [T_2(y) - T_1^2(y)/n\,\overline{s^2}]/(n-1)$是 σ^2 的一个无偏估计，并且由于 T 的完备性，可知它是 MVUE。可以将$\widetilde{g}_2$ 写作

$$\widetilde{g}_2[T(\mathbf{y})] = \frac{1}{n-1}\sum_{k=1}^{n}(y_k - \hat{\mu}s_k)^2 \triangleq \hat{\sigma}^2 \tag{4.3.19}$$

其中$\hat{\mu}$就是式(4.3.15)中建立起的 μ 的 MVUE。由于$\hat{n}_k \triangleq y_k - \hat{\mu}s_k$ 是第 k 个采样的噪声估

计，所以$\hat{\sigma}^2$ 通过 $\frac{1}{n-1}\sum_{k=1}^{n}(\hat{n}_k)^2$ 估计噪声的方差(与二阶矩对应)。需要注意的是，二阶矩的一个更自然的估计子是 $\frac{1}{n}\sum_{k=1}^{n}(\hat{n}_k)^2$ 。不过，正如从上面的分析中看到的一样，估计子 $\frac{1}{n}\sum_{k=1}^{n}(\hat{n}_k)^2$ 是有偏的。关于这一点的更进一步讨论请参见 4.4 节。 ◀

以上所描述的理论为寻找最小方差无偏估计子提供了一条途径。然而，很多模型并没有给出如命题 4.3.3 的可用框架。因此，经常需要面临这样一个问题，就是在不知道估计子的最优性的情况下，提出一个估计子并评价它的性能(例如有偏性和方差)。在这种情况下，能够评价估计子性能的标准是有用的，即对于给定模型，明确其估计子性能的基本限制是有益的。这个标准在某种程度上由下面结论给出。

命题 4.3.4 信息不等式

假设$\hat{\theta}$是族$\{P_\theta;\ \theta\in\Lambda\}$中参数 θ 的估计，并且有下列条件成立：

1) Λ 是开区间；

2) 族$\{P_\theta;\ \theta\in\Lambda\}$有相应的密度族$\{p_\theta;\ \theta\in\Lambda\}$，所有元素都有相同的支撑集；[⊖]

3) $\partial p_\theta(y)/\partial\theta$ 存在，且对所有 $\theta\in\Lambda$ 都是有限的，并且所有 y 在 p_θ 的支撑集里；

4) 对于所有 $\theta\in\Lambda$、$h(y)=\hat{\theta}(y)$ 和 $h(y)=1$，$\partial\int_\Gamma h(y)p_\theta(y)\mu(\mathrm{d}y)/\partial\theta$ 存在并且等于 $\int_\Gamma h(y)[\partial p_\theta(y)/\partial\theta]\mu(\mathrm{d}y)$ 。

那么

$$\mathrm{var}_\theta[\hat{\theta}(Y)]\geqslant\frac{\left[\frac{\partial}{\partial\theta}E_\theta\{\hat{\theta}(Y)\}\right]^2}{I_\theta}\tag{4.3.20}$$

其中

$$I_\theta\triangleq E_\theta\left\{\left(\frac{\partial}{\partial\theta}\log p_\theta(Y)\right)^2\right\}\tag{4.3.21}$$

5) $\partial^2 p_\theta(y)/\partial\theta^2$ 对所有 $\theta\in\Lambda$ 成立，y 在 p_θ 的支撑集里，并且

$$\int\frac{\partial^2}{\partial\theta^2}p_\theta(y)\mu(\mathrm{d}y)=\frac{\partial^2}{\partial\theta^2}\int p_\theta(y)\mu(\mathrm{d}y)$$

如果条件 5)成立，那么 I_θ 可以通过下式计算

$$I_\theta=-E_\theta\left\{\frac{\partial^2}{\partial\theta^2}\log p_\theta(Y)\right\}\tag{4.3.22}$$

证明 利用 Schwarz 不等式可直接得到需要证明的结论。特别地，有

$$E_\theta\{\hat{\theta}(Y)\}=\int_\Gamma\hat{\theta}(y)p_\theta(y)\mu(\mathrm{d}y)\tag{4.3.23}$$

对式(4.3.23)进行微分，并且将条件 4)代入，有

⊖ 集合$\{y\mid p_\theta(y)>0\}$对于所有 $\theta\in\Lambda$ 都相同。

$$\frac{\partial}{\partial\theta}E_\theta\{\hat{\theta}(Y)\} = \int_\Gamma \hat{\theta}(y)\frac{\partial}{\partial\theta}p_\theta(y)\mu(\mathrm{d}y)$$

同时根据条件4)还能推出

$$\int_\Gamma \frac{\partial}{\partial\theta}p_\theta(y)\mu(\mathrm{d}y) = \frac{\partial}{\partial\theta}\int_\Gamma p_\theta(y)\mu(\mathrm{d}y) = \frac{\partial}{\partial\theta}(1) = 0$$

所以可得

$$\begin{aligned}\frac{\partial}{\partial\theta}E_\theta\{\hat{\theta}(Y)\} &= \int_\Gamma(\hat{\theta}(y) - E_\theta\{\hat{\theta}(Y)\})\frac{\partial}{\partial\theta}p_\theta(y)\mu(\mathrm{d}y)\\ &= \int_\Gamma(\hat{\theta}(y) - E_\theta\{\hat{\theta}(Y)\})\left[\frac{\partial}{\partial\theta}\log p_\theta(y)\right]p_\theta(y)\mu(\mathrm{d}y) \qquad (4.3.24)\\ &= E_\theta\left\{[\hat{\theta}(y) - E_\theta\{\hat{\theta}(Y)\}]\left[\frac{\partial}{\partial\theta}\log p_\theta(Y)\right]\right\}\end{aligned}$$

其中第二个等式由 $\partial\log p_\theta(y)/\partial\theta = [\partial p_\theta(y)/\partial\theta]/p_\theta(y)$ 推得。将 Schwarz 不等式代入式(4.3.24)后，有

$$\left(\frac{\partial}{\partial\theta}E_\theta\{\hat{\theta}(Y)\}\right)^2 \leqslant E_\theta\{[\hat{\theta}(Y) - E_\theta\{\hat{\theta}(Y)\}]^2\}I_\theta \qquad (4.3.25)$$

其中 I_θ 按照式(4.3.21)定义。结合 $E_\theta\{[\hat{\theta}(Y) - E_\theta\{\hat{\theta}(Y)]^2\} = \mathrm{var}_\theta[\hat{\theta}(Y)]$ 可证明式(4.3.20)成立。

为了证明式(4.3.22)，由

$$\frac{\partial^2}{\partial\theta^2}\log p_\theta(Y) = \left(\frac{\partial^2}{\partial\theta^2}p_\theta(Y)/p_\theta(Y)\right) - \left(\frac{\partial}{\partial\theta}\log p_\theta(Y)\right)^2 \qquad (4.3.26)$$

在式(4.3.26)等号两边同时取期望 $E_\theta\{\cdot\}$ 并重新整理可得

$$I_\theta = -E_\theta\left(\frac{\partial^2}{\partial\theta^2}\log p_\theta(Y)\right) - \int_\Gamma\frac{\partial^2}{\partial\theta^2}p_\theta(y)\mu(\mathrm{d}y)$$

利用条件5)可得

$$\int_\Gamma\frac{\partial^2}{\partial\theta^2}p_\theta(y)\mu(\mathrm{d}y) = \frac{\partial^2}{\partial\theta^2}\int_\Gamma p_\theta(y)\mu(\mathrm{d}y) = \frac{\partial^2}{\partial\theta^2}(1) = 0$$

即式(4.3.22)成立。 □

由式(4.3.21)定义的量 I_θ 通常称作利用 Y 估计 θ 的 Fisher 信息，并且称式(4.2.20)为 Fisher 信息不等式。对于给定的模型，由 Fisher 信息不等式可以知道，此信息测度越高，估计精确度的下限将会越好。能够使信息不等式取等号的估计子只在某些特定的情况下才存在[参见 Lehman(1938)或下文的讨论]。对于$\hat{\theta}$是无偏估计($E_\theta\{\hat{\theta}(y)\} = \theta$)的特殊情况，Fisher 信息不等式化简为

$$\mathrm{var}_\theta[\hat{\theta}(Y)] \geqslant \frac{1}{I_\theta} \qquad (4.2.27)$$

这就是著名的克拉美罗下界(Cramér-Rao lower bound，CRLB)。

下面以特定的估计问题为例来阐释信息不等式，这些例子将进一步说明指数族在参数估计中的重要性。

例 4.3.4 指数族的信息不等式

假设 Λ 是开区间且 $p_\theta(y)$ 可表示为

$$p_\theta(y) = C(\theta)e^{g(\theta)T(y)}h(y) \tag{4.3.28}$$

其中 C、g、T 和 h 都是实值函数，且 $g(\theta)$ 存在一阶导数 $g'(\theta)$。假设 $E_\theta\{|T(Y)|\}<\infty$ 且

$$\frac{\partial}{\partial\theta}\int_\Gamma e^{g(\theta)T(y)}h(y)\mu(dy) = \int_\Gamma \frac{\partial}{\partial\theta}e^{g(\theta)T(y)}h(y)\mu(dy)$$

命题4.3.4中的条件1)～4)都成立。由于 $p_\theta(y)$ 的积分等于1，归一化后可以得到 $C(\theta) = \left[\int_\Gamma e^{g(\theta)T(y)}h(y)\mu(dy)\right]^{-1}$。

为了计算指数密度族的 Fisher 信息量 I_θ，将式(4.3.28)改写成如下形式

$$\log p_\theta(y) = g(\theta)T(y) + \log h(y) - \log\left[\int_\Gamma e^{g(\theta)T(y)}h(y)\mu(dy)\right]$$

对其进行微分后，有

$$\frac{\partial}{\partial\theta}\log p_\theta(y) = g'(\theta)T(y) - \frac{g'(\theta)\int_\Gamma T(y)e^{g(\theta)T(y)}h(y)\mu(dy)}{\int_\Gamma e^{g(\theta)T(y)}h(y)\mu(dy)} = g'(\theta)[T(y) - E_\theta\{T(Y)\}]$$

因此

$$I_\theta \triangleq E_\theta\left\{\left(\frac{\partial}{\partial\theta}\log p_\theta(Y)\right)^2\right\} = [g'(\theta)]^2 E_\theta\{[T(Y) - E_\theta\{T(Y)\}]^2\} = [g'(\theta)]^2 \text{var}_\theta[T(Y)]$$

此时 Fisher 信息不等式化为

$$\text{var}_\theta[\hat{\theta}(Y)] \geqslant \frac{\left[\frac{\partial}{\partial\theta}E_\theta\{\hat{\theta}(Y)\}\right]^2}{[g'(\theta)]^2 \text{var}_\theta[T(Y)]} \tag{4.3.29}$$

假设考虑 $T(y)$ 本身就是 θ 的估计子。那么有

$$E_\theta\{T(Y)\} = \frac{\int_\Gamma T(y)e^{g(\theta)T(y)}h(y)\mu(dy)}{\int_\Gamma e^{g(\theta)T(y)}h(y)\mu(dy)} \tag{4.3.30}$$

对式(4.3.30)进行微分后，可直接得到

$$\frac{\partial}{\partial\theta}E_\theta\{T(Y)\} = g'(\theta)\text{var}_\theta[T(Y)]$$

结合式(4.3.29)易知 Fisher 信息不等式的下界等于

$$\frac{\left[\frac{\partial}{\partial\theta}E_\theta\{\hat{\theta}(Y)\}\right]^2}{[g'(\theta)]^2 \text{var}_\theta[T(Y)]} = \text{var}_\theta[T(Y)] \tag{4.3.31}$$

从式(4.3.31)中可以看出：$T(Y)$ 达到了 Fisher 信息不等式的下界，即在所有满足 $\partial E_\theta\{\hat{\theta}(Y)\}/\partial\theta = \partial E_\theta\{T(Y)\}/\partial\theta$ 的估计子 $\hat{\theta}$ 中，它的方差是最小的。特别地，如果 T 对于 θ 是无偏的，那么它就是 MVUE。事实上，在该情况下 T 也是 θ 的完备充分统计量。◀

在上述正则化假设下，式(4.3.28)所描述的指数族是 T 的方差取得信息不等式下界的充分条件。类似地，在相同约束条件下，这种形式也是对所有 $\theta\in\Lambda$ 都取得下界的必要条件。特别值得注意的是：对于所有 $\theta\in\Lambda$ 而言，$\hat{\theta}$ 的方差取信息不等式下限的充分必要条件是式(4.3.25)取等号，即

$$\frac{\partial}{\partial\theta}\log p_\theta(Y) = k(\theta)[\hat{\theta}(Y) - E_\theta\{\hat{\theta}(Y)\}]$$

将 Λ 记为 $(a,\ b)$，$E_\theta\{\hat{\theta}(Y)\}$ 记为 $f(\theta)$，可得 $\hat{\theta}$ 取得 Fisher 信息不等式下界的充分必要条件是

$$p_\theta(y)=h(y)\exp\left\{\int_a^\theta k(\sigma)[\hat{\theta}(y)-f(\sigma)]\mathrm{d}\sigma\right\},\quad y\in\Gamma \tag{4.3.32}$$

其中 $h(y)$ 不依赖于 θ。式(4.3.32)可视为 h 已知的形如式(4.3.28)的指数族形式，即

$$C(\theta)=\exp\left\{-\int_a^\theta k(\sigma)f(\sigma)\mathrm{d}\sigma\right\},\quad g(\theta)=\int_a^\theta k(\sigma)\mathrm{d}\sigma, T(y)=\hat{\theta}(y)$$

将式(4.3.32)代入式(4.3.24)后，易知 $k(\theta)$ 必须等于 $I_\theta/[\partial E_\theta\{\hat{\theta}(\boldsymbol{Y})\}/\partial\theta]$。至此得出结论：**单参数指数族分布在正则化条件约束下，估计值 $\hat{\theta}$ 取得 Fisher 信息不等式下界的充分必要条件是 $\hat{\theta}(y)=T(y)$。**

4.4　最大似然估计

对于很多实际的观察模型，因为所要求的分析过程过于困难或是缺少可用的完备充分统计量，无法利用 4.3 节的相关结论来求解最小方差无偏估计。因此对于这些模型，有必要寻找出其他合适的估计量，例如最大似然估计。最大似然估计是常用的估算方法，下面予以重点介绍。

为了便于引出最大似然估计，首先考虑最大后验估计 $\hat{\theta}_{\mathrm{MAP}}(y)$。最大后验估计由下式给出

$$\hat{\theta}_{\mathrm{MAP}}(y)=\arg\{\max_{\theta\in\Lambda}p_\theta(y)w(\theta)\} \tag{4.4.1}$$

在缺少参数先验信息时，可以假设它在其定义域上均匀分布(即 $w(\theta)$ 是常数)，近似给出有关先验信息的最坏情况。此时，给定观察 $y\in\Gamma$ 的最大后验估计是任意一个在 Λ 上最大化 $p_\theta(y)$ 的 θ。又因为 $p_\theta(y)$ 是 θ 的函数，也称为**似然函数**($p_1(y)/p_0(y)$ 是似然比)，所以通常将等先验信息的最大后验估计称为**最大似然估计**(MLE)。用 $\hat{\theta}_{\mathrm{ML}}$ 表示这个估计值，即

$$\hat{\theta}_{\mathrm{ML}}(y)=\arg\{\max_{\theta\in\Lambda}p_\theta(y)\} \tag{4.4.2}$$

在上述论证过程中存在两个错误：一方面，无法保证总能构造出在 Λ 上的均匀分布，因为 Λ 可能无界；另一方面，假设参数服从均匀分布与假设先验信息未知或者参数是确定量存在本质区别。不过，最大似然估计在很多情况下仍然是非常有效的。在本节中，它的具体推导过程将通过其他更直接的方式展现。此外，求解使观察最有可能发生的 θ 值，就其本身而言也是合理的。

考虑到取 $p_\theta(y)$ 的最大值与取 $\log p_\theta(y)$ 的最大值等价，在假设函数足够平滑后，最大似然估计的必要条件是

$$\frac{\partial}{\partial\theta}\log p_\theta(y)\big|_{\theta=\hat{\theta}_{\mathrm{ML}}(y)}=0 \tag{4.4.3}$$

式(4.4.3)就是**似然方程**，即便它的解并不是 $p_\theta(y)$ 的极大值，但也具有一些非常有用的性质。

比如，假设需要求解克拉美罗下界，即假设 $\hat{\theta}$ 是 θ 的无偏估计，且 $\mathrm{var}_\theta[\hat{\theta}(Y)]=1/I_\theta$ (此时 $\hat{\theta}$ 是 θ 的最小方差无偏估计)。依据式(4.3.32)，可以得到 $\log p_\theta(y)$ 一定具有如下形式

$$\log p_\theta(y)=\int_a^\theta I_\sigma[\hat{\theta}(y)-\sigma]\mathrm{d}\sigma+\log h(y) \tag{4.4.4}$$

其中需要利用 $f(\theta)=\theta$ 和 $k(\theta)=I_\theta/f'(\theta)$。进而有似然方程为

$$\frac{\partial}{\partial\theta}\log p_\theta(y)\big|_{\theta=\hat{\theta}_{\mathrm{ML}}(y)}=I_\theta[\hat{\theta}(y)-\theta]\big|_{\theta=\hat{\theta}_{\mathrm{ML}}(y)}=0 \tag{4.4.5}$$

它的解为$\hat{\theta}_{\mathrm{ML}}(y)=\hat{\theta}(y)$。显然，如果$\hat{\theta}$能达到克拉美罗下界，那么它就是似然方程的解。也就是说，只有似然方程的解能达到克拉美罗下界。遗憾的是，似然方程的解并不总是都能达到克拉美罗下界，甚至也不能保证它们都是无偏估计[不过，在 $\log p_\theta$ 取得式(4.4.4)的值时可达到克拉美罗下界]。并且，当似然方程的解不符合克拉美罗下界时，可能存在其他的具有相同偏差的估计量，能获得比$\hat{\theta}_{\mathrm{ML}}$更小的估计方差。

由上述讨论可知：似然方程的解在某些情况下可作为最小方差无偏估计量。在观测空间 $\mathbf{R}^n$ 由独立同分布的 Y 所构成的情况下，似然方程的解在正则性约束下是无偏差的，且随着 $n\to\infty$逐渐趋于克拉美罗下界。在研究渐近性之前，首先用下面的两个例子来解释最大似然估计。

例 4.4.1　指数分布参数的最大似然估计

假设 $\Gamma=\mathbf{R}^n$，$\Lambda=(0,\infty)$，且 $Y_1,\cdots,Y_n$ 为独立同分布的指数分布随机变量，θ 为其分布参数，即 $p_\theta(\mathbf{y})=\prod\limits_{k=1}^{n}f_\theta(y_k)$，有

$$f_\theta(y_k)=\begin{cases}\theta\mathrm{e}^{-\theta y_k}, & y_k\geqslant 0\\ 0, & y_k<0\end{cases} \tag{4.4.6}$$

已知 $p_\theta(\mathbf{y})=\theta^n\exp\{-\theta n\bar{y}\}$，其中 $\bar{y}\triangleq\frac{1}{n}\sum\limits_{k=1}^{n}y_k$。似然方程可写为

$$\frac{\partial}{\partial\theta}\log p_\theta(\mathbf{y})\bigg|_{\theta=\hat{\theta}_{\mathrm{ML}}(y)}=\frac{n}{\theta}-n\bar{y}\bigg|_{\theta=\hat{\theta}_{\mathrm{ML}}(y)}=0 \tag{4.4.7}$$

方程的唯一解为$\hat{\theta}_{\mathrm{ML}}(\mathbf{y})=1/\bar{y}$。因为 $\partial^2\log p_\theta(\mathbf{y})/\partial\theta^2=-n^2/\theta^2<0$，则 $p_\theta(\mathbf{y})$的极大值在该点处取得。因为 $E_\theta\{Y_k\}=1/\theta$，所以 $E_\theta\{\bar{Y}\}=1/\theta$，且将 $1/\bar{y}$ 作为 θ 的估值是合理的。实际上，弱大数定理表明 $\bar{Y}\to1/\theta$ 依概率服从 P_θ，也就是 $1/\bar{Y}\to\theta$ 依概率服从 P_θ；即最大似然估计依概率收敛于实际参数，这就是它的**一致性**。最大似然估计的一致性并不只适用于指数参数分布的例子，它在一般意义下也适用。

相应的 Fisher 信息量可通过下式计算

$$I_\theta=-E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log p_\theta(\mathbf{Y})\right]=-E_\theta\{-n/\theta^2\}=n/\theta^2$$

对应的克拉美罗下界是 θ^2/n。因为 $\partial\log p_\theta(\mathbf{y})/\partial\theta$ 不满足 $k(\theta)[\hat{\theta}_{\mathrm{ML}}(\mathbf{y})-f(\theta)]$的形式，所以 Fisher 信息不等式不能取得下界。然而，可以直接计算出$\hat{\theta}_{\mathrm{ML}}$的均值和方差。特别地，利用特征函数直接就能得出样本均值 $\bar{Y}$ 的概率分布函数为

$$p_{\bar{Y}}(\bar{y})=\begin{cases}\dfrac{(n\theta)^n}{n!}\bar{y}^{n-1}\mathrm{e}^{-n\theta\bar{y}}, & \bar{y}\geqslant 0\\ 0, & \bar{y}<0\end{cases}$$

从中可以求得(当 $n>1$ 时)

$$E_\theta\{\hat{\theta}_{\mathrm{ML}}(\mathbf{Y})\} = E_\theta\left\{\frac{1}{\overline{Y}}\right\} = \frac{n\theta}{n-1} \tag{4.4.8}$$

以及(当 $n>2$ 时)

$$\mathrm{var}_\theta[\hat{\theta}_{\mathrm{ML}}(\mathbf{Y})] = \frac{\theta^2 n^2}{(n-1)^2(n-2)} \tag{4.4.9}$$

由式(4.4.8)可以得出：尽管$\hat{\theta}_{\mathrm{ML}}(\mathbf{Y})$是有偏差的，但$\lim\limits_{n\to\infty}E_\theta\{\hat{\theta}_{\mathrm{ML}}(\mathbf{Y})\}=\theta$，即它是渐近无偏的。另外，当 $n\to\infty$时，有下式成立

$$\mathrm{var}_\theta[\hat{\theta}_{\mathrm{ML}}(\mathbf{Y})]I_\theta = \frac{n^3}{(n-1)^2(n-2)} \to 1$$

因此，$\hat{\theta}_{\mathrm{ML}}$的方差逐渐趋于克拉美罗下界，即具有渐近有效性。就一般意义上的独立同分布观察而言，渐近无偏性与渐近有效性是 MLE 的两大特征。

作为这个例子的最后注解，根据命题 4.3.3，在这个模型中，$\overline{Y}$ 是完备充分统计量。并且，由式(4.4.8)得出

$$\frac{n-1}{n}\hat{\theta}_{\mathrm{ML}}(\mathbf{y}) \equiv \left(\frac{1}{n-1}\sum_{k=1}^{n} y_k\right)^{-1}$$

上式是 θ 的无偏估计量，它仅由 $\overline{Y}$ 决定。因此 θ 的最小方差无偏估计为

$$\frac{n-1}{n}\hat{\theta}_{\mathrm{ML}}(\mathbf{Y}) \triangleq \hat{\theta}_{\mathrm{MV}}(\mathbf{y})$$

从式(4.4.9)可知，估计方差为

$$\mathrm{var}_\theta[\hat{\theta}_{\mathrm{ML}}(\mathbf{Y})] = \frac{\theta^2}{(n-2)} \tag{4.4.10}$$

它是比克拉美罗下界更大的数(这是必然的，因为该情况下无法达到克拉美罗下界)，但是随着 n 的增大，逐渐接近克拉美罗下界。

式(4.4.10)中的方差等于$\hat{\theta}_{\mathrm{MV}}$的均方误差，因为它是无偏估计。对于最大似然估计，它的均方误差为

$$E_\theta\{[\hat{\theta}_{\mathrm{ML}}(\mathbf{Y})-\theta]^2\} = \mathrm{var}_\theta[\hat{\theta}_{\mathrm{ML}}(\mathbf{Y})] + b^2(\theta) \tag{4.4.11}$$

其中 $b(\theta)\triangleq E_\theta\{\hat{\theta}_{\mathrm{ML}}(\mathbf{Y})\}-\theta$ 是$\hat{\theta}_{\mathrm{ML}}$的偏差。利用式(4.4.8)和(4.4.9)可以得到

$$E_\theta\{[\hat{\theta}_{\mathrm{ML}}(\mathbf{Y})-\theta]^2\} = \frac{\theta^2(n+2)}{(n-1)(n-2)}$$

显然，$\hat{\theta}_{\mathrm{MV}}$的均方误差大于 $\theta^2/(n-2)$。因此，在这种情况下的最小方差无偏估计比最大似然估计更适合，尽管它们渐近相等。 ◀

例 4.4.2 信号幅度的最大似然估计

考虑在例 4.3.3 中已分析过的模型

$$Y_k = N_k + \mu s_k, \quad k = 1,\cdots,n$$

其中 N_1，…，N_n 是服从$\mathcal{N}(0, \sigma^2)$的独立同分布噪声，$\mathbf{s}=(s_1, \cdots, s_n)^{\mathrm{T}}$ 已知，方差 σ^2 已知时幅度 μ 的似然方程为

$$-\frac{\partial}{\partial\mu}\left(\frac{1}{2}\sum_{k=1}^{n}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\sum_{k=1}^{n}(y_k-\mu s_k)^2\right)\bigg|_{\mu=\hat{\mu}_{\mathrm{ML}}(\mathbf{y})}$$

$$= \frac{1}{\sigma^2}\sum_{k=1}^{n} s_k[y_k - \hat{\mu}_{\mathrm{ML}}(y)s_k] = 0 \tag{4.4.12}$$

这意味着

$$\hat{\mu}_{\mathrm{ML}}(\mathbf{y}) = \frac{1}{n}\sum_{k=1}^{n} s_k y_k / \overline{s^2} \tag{4.4.13}$$

其中$\overline{s^2} \triangleq (1/n)\sum_{k=1}^{n} s_k^2$。因为

$$-\frac{\partial}{\partial\mu^2}\sum_{K=1}^{n}\left(\frac{1}{2}\log(2\pi\sigma^2) + \frac{1}{2\sigma^2}(y_k - \mu s_k)^2\right) = -n\overline{s^2}/\sigma^2 < 0 \tag{4.4.14}$$

所以 $\log p_\theta(y)$是 μ 的凹函数，进而似然方程的解是全局极大值。

因为$\hat{\mu}_{\mathrm{ML}}$与 μ 的最小方差无偏估计相等(由例 4.3.3 可看出)，由此可得 $E_\theta\{\hat{\mu}_{\mathrm{ML}}(\mathbf{Y})\}=\mu$ 和 $\mathrm{var}_\theta[\hat{\mu}_{\mathrm{ML}}(\mathbf{Y})]=\sigma^2/n\overline{s^2}$。由式(4.4.14)可知 $I_\theta=n\overline{s^2}/\sigma^2$，则有

$$\mathrm{CRLB} = \frac{\sigma^2}{n\overline{s^2}} = \mathrm{var}_\theta[\hat{\mu}_{\mathrm{ML}}(\mathbf{Y})] \tag{4.4.15}$$

到当 $\theta=\mu$ 时，可以得到

$$\frac{\partial}{\partial\theta}\log p_\theta(\mathbf{y}) = k(\theta)[\hat{\theta}_{\mathrm{ML}}(\mathbf{y}) - \theta]$$

其中 $k_\theta=I_\theta=n\overline{s^2}/\sigma^2$，这正是克拉美罗下界所要求的形式。

现在假设 μ 已知，需要估算 σ^2。似然方程变为

$$\left.\frac{\partial}{\partial\sigma^2}\log p_\theta(\mathbf{y})\right|_{\sigma^2=\hat{\sigma}^2_{\mathrm{ML}}(\mathbf{y})} = \frac{1}{2\,\hat{\sigma}^2_{\mathrm{ML}}(\mathbf{y})} - \frac{1}{2[\hat{\sigma}^2_{\mathrm{ML}}(\mathbf{y})]^2}\sum_{k=1}^{n}(y_k - \mu s_k)^2 = 0 \tag{4.4.16}$$

该方程有唯一解

$$\hat{\sigma}^2_{\mathrm{ML}}(\mathbf{y}) = \frac{1}{n}\sum_{k=1}^{n}(y_k - \mu s_k)^2 \tag{4.4.17}$$

因为

$$\frac{\partial}{\partial\sigma^2}\log p_\theta(\mathbf{y}) = \frac{n}{2\sigma^4}[\hat{\sigma}^2_{\mathrm{ML}}(\mathbf{y}) - \sigma^2] \tag{4.4.18}$$

所以，当 $\sigma^2<\hat{\sigma}^2_{\mathrm{ML}}(\mathbf{y})$时，$\log p_\theta(\mathbf{y})$在 σ^2 上是增加的；当 $\sigma^2>\hat{\sigma}^2_{\mathrm{ML}}(\mathbf{y})$时，$\log p_\theta(\mathbf{y})$在 σ^2 上是减小的。显然，$\log p_\theta(\mathbf{y})$在$\hat{\sigma}^2_{\mathrm{ML}}(\mathbf{y})$处得到极大值。还可以由式(4.4.18)得到，当 $\theta=\sigma^2$ 时，有

$$\frac{\partial}{\partial\theta}\log p_\theta(\mathbf{y}) = \frac{n}{2\theta^2}[\hat{\theta}_{\mathrm{ML}}(\mathbf{y}) - \theta] \tag{4.4.19}$$

例 4.3.4 说明$\hat{\sigma}^2_{\mathrm{ML}}(\mathbf{y})$无偏且能达到克拉美罗下界，因此$\hat{\theta}^2_{\mathrm{ML}}$是 σ^2 的最小方差无偏估计。通过观察式(4.4.19)，得到 $I_\theta=n/2\theta^2\equiv n/2\sigma^4$，所以

$$\mathrm{CRLB} = \frac{2\sigma^4}{n} = \mathrm{var}_\theta[\hat{\sigma}^2_{\mathrm{ML}}(\mathbf{y})] \tag{4.4.20}$$

现在假设 μ 和 σ^2 都未知。令 $\theta=(\mu,\sigma^2)$，θ 的最大似然估计通过在 μ 和 σ^2 的值域上最大化 $p_\theta(\mathbf{y})$得到。既然从式(4.4.13)可知$\hat{\mu}_{\mathrm{ML}}(\mathbf{y})$的极大值与 σ^2 无关，可以得到

$$\max_{(\mu,\sigma^2)}\log p_\theta(\mathbf{y}) = \max_{\sigma^2}\{\max_{\mu}\log p_\theta(\mathbf{y})\} = \max_{\sigma^2}\left\{-\frac{1}{2}\sum_{k=1}^{n}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k=1}^{n}[y_k - \hat{\mu}_{\mathrm{ML}}(\mathbf{y})s_k]^2\right\}$$

除了将 μ 设为$\hat{\mu}_{\mathrm{ML}}(\mathbf{y})$之外，方程的右边与已知 μ 时估计 σ^2 属于相同的极大化问题。因此，用$\hat{\mu}_{\mathrm{ML}}$替代 μ 后，极大值直接由式(4.4.17)得到。同理，$\theta=(\mu,\sigma^2)$的最大似然估计为

$\hat{\theta}_{\mathrm{ML}}(\boldsymbol{y})=[\hat{\mu}_{\mathrm{ML}}(\boldsymbol{y}), \hat{\sigma}^2_{\mathrm{ML}}(\boldsymbol{y})]$，其中

$$\hat{\mu}_{\mathrm{ML}}(\boldsymbol{y}) = \frac{1}{n}\sum_{k=1}^{n} s_k y_k / \overline{s^2} \tag{4.4.21a}$$

和

$$\hat{\sigma}^2_{\mathrm{ML}}(\boldsymbol{y}) = \frac{1}{n}\sum_{k=1}^{n}[y_k - \hat{\mu}_{\mathrm{ML}}(\boldsymbol{y})s_k]^2 \tag{4.4.21b}$$

在这种情况下，$\hat{\mu}_{\mathrm{ML}}$的估值仍然是 μ 的最小方差无偏估计。然而，由式(4.3.19)可知 $\hat{\sigma}^2_{\mathrm{ML}}(\boldsymbol{y})=[(n-1)/n]\hat{\sigma}^2_{\mathrm{MV}}(\boldsymbol{y})$。因此

$$E_\theta\{\hat{\sigma}^2_{\mathrm{ML}}(\boldsymbol{Y})\} = \frac{n-1}{n}\sigma^2$$

此时，σ^2 的最大似然估计是有偏的(虽然它渐近无偏)。由于

$$\mathrm{var}_\theta[\hat{\sigma}^2_{\mathrm{ML}}(\boldsymbol{Y})] = [(n-1)^2/n^2]\mathrm{var}_\theta[\hat{\sigma}^2_{\mathrm{MV}}(\boldsymbol{Y})]$$

因此$\hat{\sigma}^2_{\mathrm{ML}}(\boldsymbol{y})$有比最小方差无偏估计更小的方差，可以表示为(见习题 14)[⊖]

$$\mathrm{var}_\theta(\hat{\sigma}^2_{\mathrm{MV}}(\boldsymbol{Y})) = \frac{2\sigma^4}{n-1} \tag{4.4.22}$$

即 σ^2 的最小方差无偏估计的均方误差 $E_\theta\{[\sigma^2_{\mathrm{MV}}(\boldsymbol{Y})-\sigma^2]^2\}=2\sigma^4/(n-1)$。相应地，$\sigma^2$ 的最大似然估计的均方误差可表示为

$$\begin{aligned} E_\theta\{[\sigma^2_{\mathrm{ML}}(\boldsymbol{Y})-\sigma^2]^2\} &= \mathrm{var}_\theta[\sigma^2_{\mathrm{ML}}(\boldsymbol{Y})] + [E_\theta\{\hat{\sigma}^2_{\mathrm{ML}}(\boldsymbol{Y})\}-\sigma^2]^2 \\ &= \frac{(n-1)^2}{n^2}\frac{2\sigma^4}{n-1} = \left(\frac{n-1}{n}\sigma^2-\sigma^2\right)^2 = \sigma^4\frac{(2n-1)}{n^2} \end{aligned} \tag{4.4.23}$$

这两个数值的比率为

$$\frac{E_\theta\{[\hat{\sigma}^2_{\mathrm{MV}}(\boldsymbol{Y})-\sigma^2]^2\}}{E_\theta\{[\hat{\sigma}^2_{\mathrm{MV}}(\boldsymbol{Y})-\sigma^2]^2\}} = \left(\frac{n}{n-1}\right)\left(\frac{2n}{2n-1}\right) > 1 \tag{4.4.24}$$

由式(4.4.24)可知，*均方误差一致低于最小方差无偏估计的均方误差*。这是因为最大似然估计的偏差在 MSE 意义上所产生的方差增量被最小方程无偏估计在方差上所引入的更大增量所抵消。因此，就均方误差而言，最小方差无偏估计并不一定总是最优估计。 ◀

上面的两个例子展示了最大似然估计的一个重要特点，即随着独立样本个数的增加，最大似然估计具有渐近性。这也是最大似然估计获得广泛应用的主要原因。下面将讨论最大似然估计的渐近性。

假设有一系列独立同分布采样 $Y_1, Y_2, \cdots, Y_n$，每个样本都有集合$\{f_\theta; \theta\in\Lambda\}$中的边缘密度函数 f_θ。将$\hat{\theta}_n$ 记作样本个数为 n 的似然方程的解，即

$$\frac{\partial}{\partial\theta}\log p_\theta(\boldsymbol{y})\Big|_{\theta=\hat{\theta}_n(\boldsymbol{y})} = \sum_{k=1}^{n}\psi[y_k;\hat{\theta}_n(\boldsymbol{y})] = 0$$

其中 $\psi(y_k;\theta)\triangleq\partial\log f_{\theta(y_k)}/\partial\theta$。等价为

$$\frac{1}{n}\sum_{k=1}^{n}\psi[y_k;\hat{\theta}_n(\boldsymbol{y})] = 0 \tag{4.4.25}$$

⊖ 有趣的是：μ 已知情况下 σ^2 的最小方差无偏估计的方差为 $2\sigma^4/n$[从式(4.4.20)可知]；μ 未知情况下 σ^2 的最小方差无偏估计的方差为 $2\sigma^4/(n-1)$[从式(4.4.22)可知]。因此，对于 σ^2 的无偏估计，当 μ 未知时，存在着一次观测“损失”。

下面固定参数 $\theta'\in\Lambda$，分析参量 $\sum_{k=1}^{n}\psi(y_k,\theta')/n$ 的值。假设 θ 是实际参数值，即 $Y_k\sim f_\theta$，弱大数定理表明

$$\frac{1}{n}\sum_{k=1}^{n}\psi(y_k;\theta')\xrightarrow{\text{依概率}}E_\theta\{\psi(Y_1;\theta')\}$$

可得

$$E_\theta\{\psi(Y_1;\theta')\}=\int_{\mathbf{R}}\frac{\partial}{\partial\theta}\log f_\theta(y_1)\,|_{\theta=\theta'}f_\theta(y_1)\mu(\mathrm{d}y_1)\triangleq J(\theta;\theta')\tag{4.4.26}$$

假设式(4.4.26)中积分和微分的顺序可以互换，则 $J(\theta;\ \theta')$ 可写为

$$J(\theta;\theta')=\int\left[\frac{\partial}{\partial\theta}\log f_\theta(y_1)\right]f_\theta(y_1)\mu(\mathrm{d}y_1)=\int\frac{\partial}{\partial\theta}f_\theta(y_1)\mu(\mathrm{d}y_1)=\frac{\partial}{\partial\theta}\int f_\theta(y_1)\mu(\mathrm{d}y_1)=\frac{\partial}{\partial\theta}(1)=0$$

因此，$\theta'=\theta$ 是方程 $J(\theta;\ \theta')=0$ 的一个解。假设它是 $J(\theta;\ \theta')$ 的唯一解，并且 $J(\theta;\ \theta')$ 和 $\sum_{k=1}^{n}\psi(Y_k;\theta')/n$ 都是 θ' 的平滑函数。考虑到对于较大的 n 值，$\sum_{k=1}^{n}\psi(Y_k;\theta')/n$ 接近于 $J(\theta;\ \theta')$，那么可以认为当 n 较大时，这两个函数具有相同的根。也就是说，当 n 较大时，$\hat{\theta}_n(\mathbf{Y})$ 趋近于实际参数值 θ。从统计意义上看，可以认为当 $n\to\infty$ 时，$\hat{\theta}_n(\mathbf{Y})\to\theta$。实际上，在满足合适的光滑度和单值性条件下，似然方程的解具有一致性，即它们依概率收敛于真实参数的值

$$\lim_{n\to\infty}P_\theta(|\hat{\theta}_n(\mathbf{Y})-\theta|>\varepsilon)=0,\quad 任意\ \varepsilon>0$$

下面给出似然方程的解具有一致性所要求的约束条件。

命题 4.4.1　似然估计的一致性

假设 $\{Y_k\}_{k=1}^{\infty}$ 是独立同分布随机变量序列，其概率密度函数为 f_θ，并且 J 和 ψ 的定义同上。若进一步假设下列三个条件成立：

1) $J(\theta;\ \theta')$ 是 θ' 的连续函数，且有唯一根 $\theta'=\theta$，函数在该点改变符号；

2) $\psi(Y_k;\ \theta')$ 是 θ' 的连续函数(概率为 1)；

3) 对于每个 n，$\sum_{k=1}^{n}\psi(Y_k;\theta')/n$ 都有唯一根 $\hat{\theta}_n$(概率为 1)。

那么 $\hat{\theta}_n\to\theta$(依概率)。

证明　根据条件 1)，选择 $\varepsilon>0$ 使得 $J(\theta;\ \theta+\varepsilon)$ 和 $J(\theta;\ \theta-\varepsilon)$ 的符号相反。令 $\delta=\min\{|J(\theta;\ \theta+\varepsilon)|,\ |J(\theta;\ \theta-\varepsilon)|\}$，对每个 n 定义 $A_n=A_n^+\cap A_n^-$，其中，集合 A_n^+ 和 A_n^- 分别为

$$\begin{aligned}A_n^+&=\left\{|J(\theta;\theta+\varepsilon)-\frac{1}{n}\sum_{k=1}^{n}\psi(Y_k;\theta+\varepsilon)|\leqslant\delta\right\}\\A_n^-&=\left\{|J(\theta;\theta-\varepsilon)-\frac{1}{n}\sum_{k=1}^{n}\psi(Y_k;\theta-\varepsilon)|\leqslant\delta\right\}\end{aligned}\tag{4.4.27}$$

那么，对于 A_n^+，$\sum_{k=1}^{n}\psi(Y_k;\theta+\varepsilon)/n$ 一定与 $J(\theta;\theta+\varepsilon)$ 同号；对于 A_n^-，$\sum_{k=1}^{n}\psi(Y_k;\theta-\varepsilon)/n$ 一定与 $J(\theta;\theta-\varepsilon)$ 同号；对于 A_n，$\sum_{k=1}^{n}\psi(Y_k;\theta+\varepsilon)/n$ 与 $\sum_{k=1}^{n}\psi(Y_k;\theta-\varepsilon)/n$ 异号。结合函数的连续性假设 2)，$\sum_{k=1}^{n}\psi(Y_k;\theta')/n$ 只有在过零点时改变符号。因此，对于 A_n，根 $\hat{\theta}_n$ 是介于 $\theta-\varepsilon$ 和 $\theta+\varepsilon$

之间的数。这意味着 A_n 是$\{|\hat{\theta}_n-\theta|\leqslant\varepsilon\}$ 的子集，有 $P(|\hat{\theta}_n-\theta|\leqslant\varepsilon)\geqslant P(A_n)$ 成立。

由弱大数定理可知

$$\frac{1}{n}\sum_{k=1}^{n}\psi(Y_k;\theta+\varepsilon)\rightarrow J(\theta;\theta+\varepsilon)(\text{依概率})$$

同时

$$\frac{1}{n}\sum_{k=1}^{n}\psi(Y_k;\theta-\varepsilon)\rightarrow J(\theta;\theta-\varepsilon)(\text{依概率}) \tag{4.4.28}$$

所以，当 $n\rightarrow\infty$时，有 $P(A_n^+)\rightarrow1$ 且 $P(A_n^-)\rightarrow1$。得到

$$\begin{aligned}1\geqslant &P(|\hat{\theta}_n-\theta|\leqslant\varepsilon)\geqslant P(A_n)=P(A_n^+)+P(A_n^-)-P(A_n^+\cup A_n^-)\\ \geqslant &P(A_n^+)+P(A_n^-)-1\rightarrow1\end{aligned} \tag{4.4.29}$$

因此 $P(\hat{\theta}_n-\theta|\leqslant\varepsilon)\rightarrow1$。又因为 ε 是任意选取的，所以似然估计的一致性成立。 □

备注 可以采用不同的方式来松弛命题中的约束条件。首先，函数 $J(\theta;\ \theta')$和$\psi(Y_k;\ \theta')$的连续性可以放宽到在 $\theta'=\theta$ 的临域上连续。此外，并不一定需要假设存在根$\hat{\theta}_n$，因为上述证明过程表明了似然方程在 A_n 上以概率 1 一定有解。实际上，只要在满足局部连续性假设的前提下，上面的过程就能够证明似然方程的一系列根以概率 1 收敛到$J(\theta;\ \theta')$的任意孤立根上。否则，如果 $J(\theta;\ \theta')$有多重根，求解似然方程时将出现不一致的解序列。

除了一致性外，从上面的例子中可以看到似然方程的解也具有渐近无偏性和有效性。在命题 4.4.1 的条件下，$\hat{\theta}_n$ 依概率收敛于θ。如果能够用下式表示这类收敛性

$$\lim_{n\rightarrow\infty}E_\theta\{\hat{\theta}_n\}=E_\theta\{\lim_{n\rightarrow\infty}\hat{\theta}_n\} \tag{4.4.30}$$

那么，就能证明估计子的渐近无偏性。式(4.4.30)中交换极限与期望的顺序对于依概率收敛而言并不总是可行的，不过，通过对 ψ 施加一系列约束后，就可以交换极限与期望的顺序。能够交换极限与期望顺序的一个充分条件是：存在随机变量 X，对于每个 n，$|\hat{\theta}_n|\leqslant X$ 成立，并且 $E_\theta\{X\}<\infty$。这就是控制收敛定理。鉴于$\hat{\theta}_n$ 具有一致性，渐近无偏性就是可预期的合理特性。

下面将说明最大似然估计具有渐近有效性的原因。首先考虑一个与之相关的问题，即寻找误差$(\hat{\theta}_n-\theta)$的渐近分布。

命题 4.4.2 似然估计的渐近正态性

假设$\{Y_k\}_{k=1}^{\infty}$是独立同分布随机变量序列，其概率密度函数为 f_θ，并且$\{\hat{\theta}_n\}_{n=1}^{\infty}$是似然方程的一系列一致根。进一步假设 ψ 满足以下五个正则条件：

1) $0<i_\theta\triangleq E_\theta\{[\psi(Y_1;\ \theta)]^2\}<\infty$；

2) 导数 $\psi'(Y_1;\ \theta')\triangleq\partial\psi(Y_1;\ \theta')/\partial\theta'$和 $\psi''(Y_k;\ \theta')\triangleq\partial^2\psi(Y_k;\ \theta')/(\partial\theta')^2$ 以概率 1 存在；

3) 存在函数 $M(Y_1)$使得对于所有的 $\theta'\in\Lambda$，有$|\psi''(Y_1;\ \theta')|\leqslant M(Y_1)$和 $E_\theta\{M(Y_1)\}<\infty$ 成立；

4) $J(\theta;\ \theta)=0$，其中 $J(\theta;\ \theta')$由式(4.4.26)定义；

5) 保留命题 4.3.4 的条件 5)。

那么

$$P_\theta(\sqrt{ni_\theta}(\hat{\theta}_n-\theta)\leqslant x)\rightarrow\Phi(x),\quad \text{对于所有 } x\in\mathbf{R}$$

其中，Φ 是标准高斯分布，即$\sqrt{n}(\hat{\theta}_n-\theta)$依分布收敛于正态分布$\mathcal{N}(0,\ 1/i_\theta)$。

证明　利用泰勒展开定理，可以将似然方程 $\frac{1}{n}\sum_{k=1}^{n}\psi(Y_k;\hat{\theta}_n)=0$ 的左半部分展开成关于 θ 的形式，即

$$\frac{1}{n}\sum_{k=1}^{n}\psi(Y_k;\theta)+(\hat{\theta}_n-\theta)\frac{1}{n}\sum_{k=1}^{n}\psi'(Y_k;\theta)+\frac{1}{2}(\hat{\theta}_n-\theta)^2\frac{1}{n}\sum_{k=1}^{n}\psi''(Y_k;\bar{\theta}_n)=0 \tag{4.4.31}$$

其中 $\bar{\theta}_n$ 介于 θ 与 $\hat{\theta}_n$ 之间。整理式(4.4.31)，得到一个关于 $\sqrt{n}(\hat{\theta}_n-\theta)$ 的表达式

$$\sqrt{n}(\hat{\theta}_n-\theta)=\frac{-\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\psi(Y_k;\theta)}{\frac{1}{n}\sum_{k=1}^{n}\psi'(Y_k;\theta)+(\hat{\theta}_n-\theta)\frac{1}{2n}\sum_{k=1}^{n}\psi''(Y_k;\bar{\theta}_n)} \tag{4.4.32}$$

下面分析式(4.4.32)右半部分的分母。根据弱大数定理，分母中的第一项 $\sum_{k=1}^{n}\psi'(Y_k;\theta)/n$ 依概率收敛于 $E_\theta\{\psi'(Y_1;\theta)\}$。利用条件 3)，第二项 $(\hat{\theta}_n-\theta)\sum_{k=1}^{n}\psi''(Y_k;\bar{\theta}_n)/2n$ 的边界是

$$\left|\frac{1}{2}(\hat{\theta}_n-\theta)\frac{1}{n}\sum_{k=1}^{n}\psi''(Y_k;\bar{\theta}_n)\right|\leqslant\frac{1}{2}\left|\hat{\theta}_n-\theta\right|\frac{1}{n}\sum_{k=1}^{n}M(Y_k) \tag{4.4.33}$$

那么，$|\bar{\theta}_n-\theta|\to 0$（依概率），且利用弱大数定理有 $\frac{1}{n}\sum_{k=1}^{n}M(Y_k)\to E_\theta\{M(Y_1)\}<\infty$ 成立。因此，第二项依概率收敛于零。综上，分母则依概率收敛于 $E_\theta\{\psi'(Y_1;\theta)\}$。

式(4.4.32)分子的求和项 $\sum_{k=1}^{n}\psi(Y_k;\theta)$ 是 n 个独立同分布随机变量的和，它们各自的均值为 $E_\theta\{\psi(Y_1;\theta)\}=J(\theta;\theta)=0$，方差为 $E_\theta\{\psi^2(Y_1;\theta)\}=i_\theta<\infty$。因此，根据中心极限定理，$-(1/\sqrt{n})\sum_{k=1}^{n}\psi(Y_k;\theta)$ 依分布收敛于正态分布 $\mathcal{N}(0,i_\theta)$。

上述两个结果表明 $\sqrt{n}(\hat{\theta}_n-\theta)$ 依分布收敛于正态分布 $\mathcal{N}(0,v^2)$，并且

$$v^2=i_\theta/E_\theta^2\{\psi'(Y_1;\theta)\} \tag{4.4.34}$$

除了利用式(4.3.22)的推导之外，直接代入 $E_\theta\{\psi'(Y_k;\theta)\}=-E_\theta\{\psi^2(Y_1;\theta)\}=-i_\theta$ 后，同样有 $v^2=1/i_\theta$ 成立。 □

备注　可见，在独立同分布情况下，Fisher 信息量由 $I_\theta=ni_\theta$ 给出。可以尝试将这个命题的结论作为 $\hat{\theta}_n$ 逐渐趋于正态分布 $\mathcal{N}(\theta,1/ni_\theta)$ 的条件。此时，$\hat{\theta}_n$ 的均值渐近等于 θ，方差渐近等于 $1/ni_\theta$，即克拉美罗下界。实际上，已经证明了 $\sqrt{n}(\hat{\theta}_n-\theta)$ 的渐近分布的均值是零，方差是 $1/i_\theta$。这与 $E_\theta\{\sqrt{n}(\hat{\theta}_n-\theta)\}\to 0$ 以及 $\mathrm{var}_\theta[\sqrt{n}(\hat{\theta}_n-\theta)]\to 1/ni_\theta$ 是不同的。在增加一些附加条件后，后两个条件(第二个条件为渐近有效)也成立，具体可由式(4.4.32)得到。从应用角度出发，命题 4.4.2 的结论足以证明最大似然估计是渐近最优(MVUE)估计。此外，渐近无偏性和渐近有效性也常用渐近误差分布的均值和方差来定义。

4.5　最大似然估计的进一步扩展

4.5.1　向量参数估计

事实上，前面有关最大似然估计的所有分析都是维度为 m 的向量参数估计的特例。在

向量参数情况下，似然方程具的向量形式为

$$
\begin{gathered}
\frac{\partial}{\partial \theta_1}\log p_{\boldsymbol{\theta}}(y)\mid_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0 \\
\vdots \\
\frac{\partial}{\partial \theta_m}\log p_{\boldsymbol{\theta}}(y)\mid_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0
\end{gathered}
\tag{4.5.1}
$$

对于独立同分布模型，有

$$
\begin{gathered}
\sum_{k=1}^{n}\psi_1(y_k;\hat{\boldsymbol{\theta}}_n) = 0 \\
\vdots \\
\sum_{k=1}^{n}\psi_m(y_k;\hat{\boldsymbol{\theta}}_n) = 0
\end{gathered}
\tag{4.5.2}
$$

其中 $\psi_j(y_k;\boldsymbol{\theta})=\partial\log f_{\boldsymbol{\theta}}(y_k)/\partial\theta_j$ 和 $f_{\boldsymbol{\theta}}$是 Y_k 的边缘概率密度。

在正则性约束下，Fisher 信息不等式(命题 4.3.4)也可扩展到向量情况。例如，无偏估计的克拉美罗下限是

$$
\mathrm{cov}_{\boldsymbol{\theta}}(\hat{\theta}) \geqslant I_{\boldsymbol{\theta}}^{-1} \tag{4.5.3}
$$

其中 $\mathrm{cov}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})\triangleq E_{\boldsymbol{\theta}}\{(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^{\mathrm{T}}\}$，$I_\theta$ 是 $m\times m$ 维 Fisher 信息矩阵，它的第 $j-l$ 号元素为

$$
(I_{\boldsymbol{\theta}})_{j,l} = E_{\boldsymbol{\theta}}\left(\left[\frac{\partial}{\partial\theta_j}\log p_{\boldsymbol{\theta}}(Y)\right]\left[\frac{\partial}{\partial\theta_l}\log p_{\boldsymbol{\theta}}(Y)\right]\right) \tag{4.5.4}
$$

由于 $I_{\boldsymbol{\theta}}$是零均值向量的协方差矩阵，即

$$
\left(\frac{\partial}{\partial\theta_1}\log p_{\boldsymbol{\theta}}(Y),\frac{\partial}{\partial\theta_2}\log p_{\boldsymbol{\theta}}(Y),\cdots,\frac{\partial}{\partial\theta_m}\log p_{\boldsymbol{\theta}}(Y)\right)^{\mathrm{T}}
$$

因此，它至少是非负定矩阵。式(4.5.3)假设此矩阵为正定矩阵，矩阵不等式 $A\geqslant B$ 表明 $A-B$ 是半正定矩阵。对独立同分布而言，式(4.5.3)即为

$$
\mathrm{cov}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) \geqslant \frac{1}{n}\boldsymbol{i}_{\boldsymbol{\theta}}^{-1} \tag{4.5.5}
$$

其中

$$
(i_{\boldsymbol{\theta}})_{j,l} = E_{\boldsymbol{\theta}}\{\psi_j(Y_1;\boldsymbol{\theta})\psi_l(Y_1;\boldsymbol{\theta})\} \tag{4.5.6}
$$

由于约束条件与命题 4.4.1 相似，似然方程的解也具有一致性，即

$$
\|\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}\| \triangleq \left[\frac{1}{m}\sum_{j=1}^{m}[(\hat{\boldsymbol{\theta}}_n)_j-\theta_j]^2\right]^{1/2} \to 0 \quad \text{(依概率)} \tag{4.5.7}
$$

同理，约束条件与命题 4.4.2 类似，估计误差依分布满足

$$
\sqrt{n}(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}) \to \mathcal{N}(0,\boldsymbol{i}_{\boldsymbol{\theta}}^{-1}) \tag{4.5.8}
$$

显然，向量参数的结论与标量参数相似。

针对独立同分布模型，有关最大似然估计更详细的性质可以参考 Lehmann(1983)所写的书。

4.5.2 信号参数估计

最大似然估计的渐近性质也可以扩展到某些时变模型，例如实值观测模型

$$Y_k = s_k(\theta) + N_k, \quad k = 1, \cdots, n \tag{4.5.9}$$

其中$\{s_k(\theta)\}_{k=1}^n$表示信号序列，它是未知参数θ的某个已知函数，$\{N_k\}_{k=1}^n$是边缘概率密度为f的独立同分布噪声序列。为简便起见，假设θ是在区间Λ上的标量参数。

在式(4.5.9)中θ的最大似然估计通过求解下面的方程得到

$$\hat{\theta}_n = \arg\max_{\theta\in\Lambda}\Big[\sum_{k=1}^n \log f[Y_k - s_k(\theta)]\Big]$$

或者通过求解下面的等价方程得到

$$\hat{\theta}_n = \arg\min_{\theta\in\Lambda}\Big[-\sum_{k=1}^n \log f[Y_k - s_k(\theta)]\Big] \tag{4.5.10}$$

相应的似然方程为

$$\sum_{k=1}^n s'_k(\hat{\theta}_n)\varphi[Y_k - s_k(\hat{\theta}_n)] = 0 \tag{4.5.11}$$

其中，$\psi \triangleq -f'/f$，$f'(x) \triangleq \mathrm{d}f(x)/\mathrm{d}x$，$s'_k(\hat{\theta}_n) \triangleq \partial s_k(\theta)/\partial\theta$。例如，当$f$是正态分布$\mathcal{N}(0, \sigma^2)$时，式(4.5.10)和式(4.5.11)分别等于

$$\hat{\theta}_n = \arg\Big[\min_{\theta\in\Lambda}\sum_{k=1}^n [Y_k - s_k(\hat{\theta}_n)]^2\Big] \tag{4.5.12}$$

和

$$\sum_{k=1}^n s'_k(\hat{\theta}_n)[Y_k - s_k(\hat{\theta}_n)] = 0 \tag{4.5.13}$$

在式(4.5.12)中，如果选择θ使得$\{s_k(\theta)\}_{k=1}^n$是数据的最小二乘估计，那么该式也是θ的最小二乘估计。即通过选择θ使得数据和由θ所产生的信号之间的平方误差最小。最小二乘法是一种经典估计方法，即使误差不满足高斯分布，在类似式(4.5.9)所描述的观测模型中也常常用到。

似然方程(4.5.11)的解与独立同分布下的最大似然估计具有相同的渐近性质。然而，在分析渐近性时，需要特别考虑信号的时变性。例如，对于有限个样本，如果信号幅度恒等于零(或独立于参数θ)，则不可能期望该观测模型仍满足一致性。为了说明对信号施加何种约束才能保证似然方程(4.5.11)的解与独立同分布情况下似然方程的解具有相同性质，以式(4.5.13)的最小二乘估计为例来进行分析。当ψ充分正则时，对于式(4.5.11)所描述的更一般性情况，类似的结论也成立。

结合观测模型(4.5.9)，代入最小二乘解后，方程(4.5.13)可改写为

$$\sum_{k=1}^n s'_k(\hat{\theta}_n)N_k + \sum_{k=1}^n s'_k(\hat{\theta}_n)[s_k(\theta) - s_k(\hat{\theta}_n)] = 0 \tag{4.5.14}$$

为了分析最小二乘解$\hat{\theta}_n$所需的特性，考虑每个$\theta' \in \Lambda$所对应的随机变量序列

$$J_n(\theta;\theta') \triangleq \sum_{k=1}^n s'_k(\theta')N_k + \sum_{k=1}^n s'_k(\theta')[s_k(\theta) - s_k(\theta')] \tag{4.5.15}$$

注意，在没有噪声($N_k \equiv 0$)时，$\hat{\theta}_n = \theta$是似然方程(4.5.14)的解。然而，除非$\theta' = \theta$是下面方程的唯一根，否则即使无噪声，方程(4.5.14)也无法得到正确的估计结果

$$K_n(\theta;\theta') \triangleq \sum_{k=1}^{n} s_k'(\theta')[s_k(\theta) - s_k(\theta')] \tag{4.5.16}$$

因此，为了确保似然方程的解具有一致性，要求噪声项 $\sum_{k=1}^{n} s_k'(\theta')N_k$ 是渐近可忽略的，且 $K_n(\theta;\theta')$ 项渐近存在唯一根。两边同除 $d_n > 0$ 后，式(4.5.14)的结果不变。换言之，如果可以找到一个序列 $\{d_n\}_{n=1}^{\infty}$，使得 $\sum_{k=1}^{n} s'_k(\theta')N_k/d_n$ 可渐近忽略不计，且 $K_n(\theta;\theta')/d_n$ 存在渐近的唯一根，那么式(4.5.14)的根与独立同分布情况具有相同的特性。

注意，在假设噪声服从$\mathcal{N}(0;\sigma^2)$分布时，有下式成立

$$\frac{1}{d_n}J_n(\theta;\theta') \sim N\left(\frac{1}{d_n}K_n(\theta;\theta'), \frac{\sigma^2}{d_n^2}\sum_{k=1}^{n} s_k'(\theta')^2\right) \tag{4.5.17}$$

即对于给定的 θ 和 $\theta' \in \Lambda$，当且仅当下式成立时，$J_n(\theta;\theta')/d_n$ 依概率收敛到常数

$$\lim_{n\to\infty} \frac{1}{d_n^2} \sum_{k=1}^{n} [s_k'(\theta')]^2 = 0 \tag{4.5.18}$$

和

$$\lim_{n\to\infty} \frac{1}{d_n} K_n(\theta;\theta') \tag{4.5.19}$$

利用该结论，可以证明下面的命题 4.5.1(类似命题 4.4.1)。

命题 4.5.1 最小二乘的一致性

假设由式(4.5.9)所表示的观察模型包含噪声$\mathcal{N}(0, \sigma^2)$，并存在标量序列$\{d_n\}_{n=1}^{\infty}$使得式(4.5.18)和式(4.5.19)对所有的 $\theta' \in \Lambda$ 成立。进一步假设 $s_k(\theta')$、$s_k'(\theta')$和 $J_n(\theta;\theta')$ 都是 θ' 的连续函数，且 $J(\theta;\theta')$在 $\theta'=\theta$ 存在唯一根。其中

$$J_n(\theta;\theta') \triangleq \lim_{n\to\infty} \frac{1}{d_n} K_n(\theta;\theta') \tag{4.5.20}$$

似然方程(4.5.13)存在一系列以概率 1 收敛于 θ 的根。特别地，如果对于每个 n，式(4.5.13)都有唯一根$\hat{\theta}_n$，则$\hat{\theta}_n \to \theta$(依概率)。

证明该结论与证明命题 4.4.1 几乎相同，因此留作练习。下面考虑估计信号振幅的问题(参见例 4.4.2)，其中对于已知序列$\{s_k\}_{k=1}^{\infty}$，有

$$s_k(\theta) = \theta s_k, \quad k = 1,2,\cdots,n \tag{4.5.21}$$

此时，存在 $s_k'(\theta) = s_k$ 使得 $\sum_{k=1}^{n}[s_k'(\theta)]^2 = \sum_{k=1}^{n} s_k^2$ 和 $K_n(\theta;\theta') = (\theta-\theta')\sum_{k=1}^{n} s_k^2$ 。因此，满足命题一致性的充分条件是存在序列$\{d_n\}_{n=1}^{\infty}$使得

$$0 < \lim_{n\to\infty} \frac{1}{d_n} \sum_{k=1}^{n} s_k^2 < \infty \tag{4.5.22}$$

在信号序列满足正则性约束条件下，式(4.5.9)的最小二乘估计具有渐近正态性。需要注意的是，如果 $s_k(\theta)$存在三阶微分，似然方程可以展开为关于 θ 的泰勒级数，即

$$\sum_{k=1}^{n} s_k'(\theta)[Y_k - s_k(\theta)] + (\hat{\theta}_n - \theta)\sum_{k=1}^{n}[s_k''(\theta)[Y_k - s_k(\theta)] - [s_k'(\theta)]^2]$$
$$+ \frac{1}{2}(\hat{\theta}_n - \theta)^2 \sum_{k=1}^{n}[s'''_k(\bar{\theta}_n)[Y_k - s_k(\bar{\theta}_n)] - 3s_k''(\bar{\theta}_n)s_k'(\bar{\theta}_n)] = 0 \tag{4.5.23}$$

其中 $\bar{\theta}_n$ 在 θ 和 $\hat{\theta}_n$ 之间。整理方程后，可得到

$$\hat{\theta}_n-\theta=\frac{-\sum_{k=1}^{n}s_k'(\theta)N_k}{\sum_{k=1}^{n}s_k''(\theta)N_k-\sum_{k=1}^{n}[s_k'(\theta)]^2+\frac{1}{2}(\hat{\theta}_n-\theta)\sum_{k=1}^{n}Z_k(\bar{\theta}_n)} \tag{4.5.24}$$

其中

$$Z_k(\theta')\triangleq[s_k'''(\theta')[N_k+s_k(\theta)-s_k(\theta')]-3s_k''(\theta')s_k'(\theta')]$$

根据误差的上述表达式，可以证明最小二乘估计具有渐近正态性。

命题 4.5.2　最小二乘的渐近正态性

假设式(4.5.9)表示的观测模型包含噪声 $\mathcal{N}(0,\sigma^2)$，且 $\{\hat{\theta}_n\}_{n=1}^{\infty}$ 是 θ 的最小二乘估计序列，进一步假设以下正则化条件成立：

1) 在 θ' 的值域上，存在函数 M 使得 $|Z_k(\theta')|\leqslant M(N_k)$ 成立，且 $E_\theta\{M(N_k)\}<\infty$，同时假设 $s_k(\theta)$ 的导数存在；

2) $\lim\limits_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}[s_k'(\theta)]^2>0$；

3) $\lim\limits_{n\to\infty}\sum_{k=1}^{n}[s_k''(\theta)]^2\Big/\Big[\sum_{k=1}^{n}[s_k'(\theta)]^2\Big]^2=0$。

进而，下式依分布成立

$$\left(\sum_{k=1}^{n}[s_k'(\theta)]^2\right)^{1/2}(\hat{\theta}_n-\theta)\to\mathcal{N}(0,\sigma^2) \tag{4.5.25}$$

证明该结论的过程与独立同分布情况下的证明过程类似，因此留作练习。需要注意的是，此时 Fisher 信息为

$$I_\theta=\sum_{k=1}^{n}[s_k'(\theta)]^2/\sigma^2 \tag{4.5.26}$$

因此，与独立同分布情况类似，对式(4.5.9)所描述的观测模型而言，最小二乘估计是误差服从 $\mathcal{N}(0,\sigma^2)$ 分布的渐近有效估计。

最小二乘估计的渐近正态性在估计信号幅度($s_k(\theta)=\theta_{s_k}$)上表现得更为直观。此时，因为 $Z_k(\theta')\equiv 0$ 且 $s_k''(\theta)=0$，所以渐近正态需要的唯一条件是 $\lim\limits_{n\to\infty}\sum_{k=1}^{n}s_k^2/n>0$。事实上，参见例 4.4.2，采用直接分析的方式也能得出 MLE 在这种特殊情况下所具有的性质。

下面将给出一个稍微复杂的例子。

例 4.5.1　一阶线性系统辨识

参数估计的一类重要应用是系统辨识，希望通过设置一个输入并观察其输出来推断系统的输入/输出结构。最简单的辨识问题是确定一阶时不变线性稳定系统。这类系统可用信号模型描述为

$$s_k(\theta)=\theta s_{k-1}(\theta)+u_k,\quad k=1,2,\cdots,n \tag{4.5.27}$$

其中，$|\theta|<1$，且 $\{u_k\}_{k=1}^{n}$ 是已知的输入序列。参数 θ 是齐次方程 $s_k(\theta)=\theta s_{k-1}(\theta)$ 的系数，即在线性、时不变和一阶假设下，系统特性完全由该参数确定。测量系统的输出通常受噪声影响，在假设噪声满足独立同分布时，估计 θ 与式(4.5.9)所描述的模型类似。在误差

服从$\mathcal{N}(0, \sigma^2)$分布时，考虑θ的均方误差估计结果。

假设系统(4.5.27)的初始状态$s_0(\theta)=0$，式(4.5.27)的解为

$$s_k(\theta)=\sum_{l=1}^{k}\theta^{k-1}u_l \tag{4.5.28}$$

当$n\to\infty$时，能否正确估计θ依赖于输入序列$\{u_k\}_{k=1}^{n}$。对于输入信号是常数的情况，即$u_k=1$且$k\geqslant 1$，系统的输出为

$$s_k(\theta)=\sum_{i=1}^{k}\theta^{i-1}=\sum_{m=0}^{k-1}\theta^m=\frac{1-\theta^k}{1-\theta}, s_k'(\theta)=\frac{(1-\theta^k)-k\theta^{k-1}(1-\theta)}{(1-\theta)^2}$$

这意味着

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}[s_k'(\theta)]^2=\frac{(2-\theta)^2}{(1-\theta)^4} \tag{4.5.29}$$

且

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}[s_k'(\theta')[s_k(\theta)-s_k(\theta')]]=\frac{(2-\theta')(\theta-\theta')}{(1-\theta')^2(1-\theta)} \tag{4.5.30}$$

显然，式(4.5.30)在$\theta'=\theta$处有唯一根，且取值在$|\theta'|<1$上连续。式(4.5.29)和式(4.5.30)同时意味着在$d_n=n$时命题4.5.1的假设成立。因此，似然方程具有一致的根序列。[实际上，$J(\theta;\theta')$在$\theta'=\theta$的邻域外有界且不等于零，似然方程任意根序列均具有一致性。]

此时，理解θ存在一致估计并不困难。注意到$s_k(\theta)$的渐近值是$1/(1-\theta)$，则系统对参数θ有唯一稳态解。既然通过增长观测时间能平均掉噪声，那么就能利用$k=1, 2, \cdots, \infty$时的系统输出来确定未知的系统参数。对于输入信号仅持续有限长时间的情况，稳定系统意味着稳态输出等于零。此时，命题4.5.1的假设不再成立。如果不存在测量噪声，那么利用瞬态响应可准确估计系统参数。不过，实际测量总存在噪声，因此有必要利用稳定状态下的输出来估计系统参数。在系统辨识中，通常将能够产生非零稳定输出的输入称为系统的持续激励。在线性系统识别问题中，了解与持续激励有关的特性就足够了。持续激励意味着输入信号的频率成分足够充分，以至于它能够激发系统的所有振荡模式。

对于常数输入信号，因为在$\Lambda=(-1, 1)$上$Z_k(\theta')$不是一致有界，所以无法直接应用命题4.5.2的结论。然而，如果假设θ不等于1且有界，即取$\Lambda=(-1, \theta_u)$和$\theta_u<1$，那么命题4.4.4的正则条件成立，似然方程的根的渐近正态性和一致有效性仍然成立。也就是说，$\sqrt{n}(\hat{\theta}_n-\theta)$的渐近方差是$\sigma^2(1-\theta)^4/(2-\theta)^2$。 ◀

就信号参数估计而言，最大似然估计和最小二乘估计所涉及的其他有关问题将在本章的后续部分和第7章中详细分析。在讨论下个问题之前，需要明确由命题4.5.1和4.5.2所给出的最小二乘估计的性质在一般意义上也成立。特别地，有下面的性质成立。

命题 4.5.3 非高斯噪声下最小二乘估计的一致性和渐近正态性

用$E\{N_k\}=0$和$E\{N_k^2\}=\sigma^2<\infty$代替$N_k\sim\mathcal{N}(0, \sigma^2)$后，命题4.5.1和命题4.5.2仍然成立。

不过，值得注意的是：在非高斯噪声背景下，Fisher信息矩阵不再具有式(4.5.26)所给出的形式。换言之，在非高斯噪声下，无法证明最小二乘估计具有渐近有效性。

4.5.3 信号参数的稳健估计

重新考虑式(4.5.9)所描述的参数估计问题。已经证明了最大似然估计在最小化渐近方差意义上是渐近最优的估计子。不过，正如 3.5 节所述，式(4.5.9)所涉及的统计模型在实际中只是近似有效。那么，针对特殊模型所设计的参数估计步骤是否稳健？即参数估计的性能是否对模型的细微变化敏感？

例如，考虑噪声服从$\mathcal{N}(0,1)$分布的模型。在正则性约束下，假设 $e_\theta \triangleq \lim\limits_{n\to\infty}\sum\limits_{k=1}^{n}[s_k'(\theta)]^2/n$ 存在且为正，最小二乘估计的渐近分布为$\mathcal{N}(\theta, 1/ne_\theta)$。假设噪声的实际统计特性近似服从$\mathcal{N}(0,1)$的正态分布，噪声的概率密度函数为

$$f(x)=(1-\varepsilon)\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}+\varepsilon h(x),\quad x\in\mathbf{R} \tag{4.5.31}$$

其中 $h(x)$是关于零对称的任意密度函数，其方差有限但无界，即

$$\sigma_h^2 \triangleq \int_{-\infty}^{\infty} x^2 h(x)\mathrm{d}x$$

那么，由命题 4.5.3 可知，最小二乘估计的渐近方差为

$$v_h^2 \sim \frac{(1-\varepsilon)+\varepsilon\sigma_h^2}{ne_\theta} \tag{4.5.32}$$

既然 σ_h^2 无界，对于任意 $\varepsilon>0$，v_h^2 可以无限大。换言之，最大渐近方差为

$$\sup_h[(1-\varepsilon)+\varepsilon\sigma_h^2]=\infty \tag{4.5.33}$$

显然，如果测量数据包含部分方差较大的噪声样本，则最小二乘法估计缺乏稳健性。这一情况在实际中可能发生，例如在雷达测量中，观测值中可能存在部分方差非常大而 ε 较小的脉冲干扰。在给定模型下，这种不太可能出现的较大观测值一般称为“野值”。与分析 3.5 节中的信号检测问题类似，需要在噪声的名义分布模型基础上采用其他的渐近方差准则来设计估计器。

假设式(4.5.9)中的噪声概率密度函数 f 偶对称，考虑 θ 的估计为

$$\sum_{k=1}^{n} s_k'(\hat{\theta}_n)\psi[Y_k-s_k(\hat{\theta}_n)]=0 \tag{4.5.34}$$

其中 ψ 是广义奇对称函数。当 $\psi(x)=x$ 时，式(4.5.34)就是最小二乘估计；当 $\psi(x)=-f'(x)/f(x)$时，式(4.5.34)对应最大似然估计。通常称形如式(4.5.34)的估计为 M-估计。假设$0<e_\theta<\infty$，φ、f 和$\{s_k(\theta)\}_{k=1}^{\infty}$均满足正则性，M-估计具有一致性，且渐近服从$\mathcal{N}[\theta, V(\psi,f)/ne_\theta]$分布，其中

$$V(\psi,f)\triangleq\frac{\int\psi^2 f}{\left(\int\psi' f\right)^2} \tag{4.5.35}$$

其中

$$\psi'(x)=-\mathrm{d}\psi(x)/\mathrm{d}x$$

显然，对于不确定的噪声概率密度函数类 F，通过最小化 M-估计方差可获得参数的稳健估计，即优化问题可描述为

$$\min_{\psi}\sup_{f\in\mathcal{F}} V(\psi,f) \tag{4.5.36}$$

就一般性的密度函数集 F 而言，Huber 在 1981 年首次给出了式(4.5.36)的求解方法，具体求解过程大致如下。

首先，记

$$I(f) \triangleq \int (f')^2/f \tag{4.5.37}$$

然后，令 f_{L} 是集合 F 中最小化 $I(f)$ 的密度函数，即

$$I(f_{\mathrm{L}}) = \min_{f\in F} I(f) \tag{4.5.38}$$

那么，函数 $\psi_{\mathrm{R}}(x) = -f'_{\mathrm{L}}(x)/f_{\mathrm{L}}(x)$ 就是式(4.5.36)的解。需要注意的是，对于任意 f，都有

$$V(\psi,f)\,|_{\psi=-f'f} = 1/I(f) \tag{4.5.39}$$

即给定 f 后，$[ne_\theta I(f)]^{-1}$ 是 MLE 的渐近方差，对应的 Fisher 信息矩阵为 $ne_\theta I(f)$。既然 f_{L} 是集合 F 中的函数，与之对应的最优估计(MLE)在所有估计中能最小化最坏性能。换言之，稳健 M-估计可理解为在最不利密度函数 f_{L} 上的最优估计。

针对某些不确定的模型集 F，已经得到了优化问题 $\min\limits_{f\in F} I(f)$ 的解的形式[参见 Huber (1981)]。例如，就式(4.5.31)所描述的 ε 污染 $\mathcal{N}(0,1)$ 模型而言，最不利密度函数为

$$f_{\mathrm{L}}(x) = \begin{cases} (1-\varepsilon)\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}, & |x|\leqslant k' \\ (1-\varepsilon)\mathrm{e}^{-k'(|x|-k')}\dfrac{1}{\sqrt{2\pi}}\mathrm{e}^{-(k')^2/2}, & |x|> k' \end{cases} \tag{4.5.40}$$

其中常数 k' 是下式的解

$$(1-\varepsilon)^{-1} = 2\Phi(k') - 1 + \frac{1}{k'}\left(\frac{2}{\pi}\right)^{1/2}\mathrm{e}^{-(k')^2/2} \tag{4.5.41}$$

相应的 ψ 函数为

$$\psi_k(x) = \begin{cases} x, & |x|\leqslant k' \\ k'\operatorname{sgn}(x), & |x|> k' \end{cases} \tag{4.5.42}$$

与假设检验类似，通过削弱奇异点的影响来实现信号参数的稳健估计。

有关 M-估计器和其他稳健估计方法的进一步讨论，读者除了参考 Kassam 和 Poor (1985)的综述文章之外，还可从 Huber(1981)和 Hampel(1986)的书中找到更完整的论述。

4.5.4 递归参数估计

最大似然估计在大量训练样本下具有很好的估计性能，但计算复杂。例如，在 n 个服从密度函数为 f_θ 的独立同分布样本基础上计算最大似然估计需要求解函数 $\sum\limits_{k=1}^{n}\log f_\theta(y_k)$ 的最大值。除非最大化 θ 可写出 $\boldsymbol{y}$ 的闭式函数形式，否则必须采用迭代方法寻找 $\hat{\theta}_{\mathrm{ML}}(y)$。当不存在低维的充分统计量时，迭代算法需要存储并同时处理 n 个样本。特别是随着样本数的增加，迭代算法几乎不具有可实现性。因此，有必要采用递归或序贯算法来降低最大似然估计的计算量和存储空间。

假设一致序列 $\{\hat{\theta}_n\}_{n=1}^{\infty}$ 是似然方程的解，即

$$\sum_{k=1}^{n}\psi(Y_k;\hat{\theta}_n)=0 \tag{4.5.43}$$

其中 $\psi(Y_k;\ \theta)=\partial\log f_\theta(Y_k)/\partial\theta$。因为$\{\hat{\theta}_n\}_{n=1}^{\infty}$是一致序列，当 $n\to\infty$时，序列之差$\hat{\theta}_n-\hat{\theta}_{n-1}$收敛到零，进而可将式(4.5.43)在$\hat{\theta}_{n-1}$处展开为

$$\sum_{k=1}^{n}\psi(Y_k;\hat{\theta}_{n-1})+(\hat{\theta}_n-\hat{\theta}_{n-1})\sum_{k=1}^{n}\psi'(Y_k;\hat{\theta}_{n-1})\sim 0 \tag{4.5.44}$$

其中 $\psi'(Y_k;\ \theta)=\partial\psi(Y_k;\ \theta)/\partial\theta$。重新整理式(4.5.44)后，有

$$\hat{\theta}_n\sim\hat{\theta}_{n-1}-\frac{\sum_{k=1}^{n}\psi(Y_k;\hat{\theta}_{n-1})}{\sum_{k=1}^{n}\psi'(Y_k;\hat{\theta}_{n-1})} \tag{4.5.45}$$

由于$\hat{\theta}_{n-1}$是 $\sum_{k=1}^{n-1}\psi(Y_k;\hat{\theta}_{n-1})$ 的解，式(4.5.45)右侧的分子只有 $\psi(Y_k;\ \hat{\theta}_{n-1})$一项。将分母中的求和写成

$$n\left[\frac{1}{n}\sum_{k=1}^{n}\psi'(Y_k;\hat{\theta}_{n-1})\right] \tag{4.5.46}$$

根据弱大数定理，有

$$-\frac{1}{n}\sum_{k=1}^{n}\psi'(Y_k;\theta)\to i_\theta(\text{依概率})$$

其中 $i_\theta=-E_\theta\{Y_k;\ \theta\}=E\{\psi^2(Y_k;\ \theta)\}$是每个样本的 Fisher 信息。又因为$\hat{\theta}_{n-1}\to\theta$，所以有下面的表达式成立

$$\frac{1}{n}\sum_{k=1}^{n}\psi'(Y_k;\hat{\theta}_{n-1})\to i_{\hat{\theta}_{n-1}} \tag{4.5.47}$$

综合式(4.5.45)和式(4.5.47)后，似然方程的一致解序列可近似表示为

$$\hat{\theta}_n\sim\hat{\theta}_{n-1}+\frac{\psi(Y_n;\hat{\theta}_{n-1})}{ni_{\hat{\theta}_{n-1}}} \tag{4.5.48}$$

至此，得到了$\hat{\theta}_n$ 的渐近递归方程，即仅需要利用$\hat{\theta}_{n-1}$和 Y_n 就可计算出$\hat{\theta}_n$。

对任意初值$\hat{\theta}_0$，式(4.5.48)所描述的递归方程可写成

$$\hat{\theta}_n=\hat{\theta}_{n-1}+\frac{\psi(Y_n;\hat{\theta}_{n-1})}{ni_{\hat{\theta}_{n-1}}},\quad n=1,2,\cdots \tag{4.5.49}$$

在正则性约束下，它与最大似然估计具有相同的渐近性能，包括渐近一致性和渐近有效性。事实上，上述递归算法是基于随机近似的一类广义递归参数估计算法的特例。实际应用中，递归算法在实时参数估计和在线运算方面具有较大优势。对其进行适当修正后，也能应用到缓变参数信号的实时跟踪领域。对递归算法感兴趣的读者可以参考 Nevel'son 和 Has'minskii (1973)的专著。针对形如式(4.5.9)所描述的时变问题，改进的最大似然和最小二乘递归估计算法可供利用，具体请参考 Ljung 和 Soderstrom(1982)，以及 Goodwin 和 Sin(1984)。

4.6　习题

1. 假设随机参数 Θ，对于给定的 $\Theta=\theta$，实观测 Y 服从以下分布

$$p_\theta(y)=(\theta/2)e^{-\theta|y|},\quad y\in\mathbf{R}$$

进一步假设 Θ 有先验概率

$$w(\theta)=\begin{cases}1/\theta, & 1\leqslant\theta\leqslant \mathrm{e}\\ 0, & 其他\end{cases}$$

(a) 计算观测 Y 下 Θ 的 MAP 估计；

(b) 计算观测 Y 下 Θ 的 MMSE 估计。

2. 假设实观测 Y 为

$$Y=N+\Theta S$$

其中，$N\sim\mathcal{N}(0,1)$，$P(S=1)=P(S=-1)=1/2$，Θ 的概率密度函数为

$$w(\theta)=\begin{cases}K\mathrm{e}^{\theta^2/2}, & 0\leqslant\theta\leqslant 1\\ 0, & 其他\end{cases}$$

其中，$K=\left[\int_0^1 \mathrm{e}^{\theta^2/2}\mathrm{d}\theta\right]^{-1}$。假设 N、Θ 和 S 是独立的。

(a) 计算观测 $Y=y$ 时 Θ 的 MAP 估计；

(b) 计算观测 $Y=y$ 时 Θ 的 MMSE 估计。

3. 假设 Θ 是一个随机参数，其先验概率为

$$w(\theta)=\begin{cases}\alpha\mathrm{e}^{-\alpha\theta}, & \theta\geqslant 0\\ 0, & \theta<0\end{cases}$$

其中，已知 $\alpha>0$。假设观测量 Y 是速率为 Θ 的泊松分布随机变量

$$p_\theta(y)\equiv P(Y=y|\Theta=\theta)=\frac{\theta^y\mathrm{e}^{-\theta}}{y!},y=0,1,2,\cdots$$

计算在观测量 Y 下的 MMSE 和 MAP，并说明 MMES 的计算方法。

4. 假设有一个由随机变量 Y 组成的单一观测 Y

$$Y=N+\Theta$$

其中，N 为均值为 0、方差为 σ^2 的高斯分布随机变量。参数 Θ 是独立于 N 的随机变量，其概率团函数为

$$w(\theta)=P(\Theta=\theta)=\begin{cases}1/2, & \theta=-1\\ 1/2, & \theta=+1\end{cases}$$

(a) 计算 $\hat{\theta}_{\mathrm{MMSE}}$ 和 $\hat{\theta}_{\mathrm{MAP}}$(需要考虑参数集 Λ 属于 $\mathbf{R}$)。

(b) 在什么条件下(a)中的两个估计值近似相等？

5. 假设 Θ 是一个随机参数，其先验概率为

$$w(\theta)=\begin{cases}\mathrm{e}^{-\theta}, & \theta\geqslant 0\\ 0, & \theta<0\end{cases}$$

而且 Y 的条件密度为

$$p_\theta(y)=\frac{1}{2}\mathrm{e}^{-|y-\theta|},\quad -\infty<y<\infty$$

计算 $\hat{\theta}_{\mathrm{MMSE}}$ 和 $\hat{\theta}_{\mathrm{MAP}}$。

6. 假设 N_1 和 N_2 是两个联合高斯随机变量，均值为 0，方差为 1，相关系数为 $\rho(|\rho|<1)$。假设观测量 Y_1 和 Y_2 为

$$Y_k=\frac{N_k}{\sqrt{\Theta}},\quad k=1,2$$

其中，Θ 为独立于 N_1 和 N_2 的随机参数，其先验密度函数为

$$\omega(\theta)=\begin{cases}1/\alpha, & \theta\in[0,\alpha]\\ 0, & \theta\notin[0,\alpha]\end{cases}$$

其中，已知 $\alpha>0$。

(a) 计算 Θ 的最小均方误差估计。

(b) 计算 Θ 的 MAP 估计。

(c) 计算 Θ 的最小绝对均值误差估计。

7. 假设 Θ 均匀分布在(0，1)区间，观测量 $Y=N+\Theta$，其中 N 独立于 Θ 的随机变量，其密度函数为

$$p_N(n)=\begin{cases}\mathrm{e}^{-n}, & n\geqslant 0\\ 0, & n<0\end{cases}$$

计算 $\hat{\theta}_{\mathrm{MMSE}}$、$\hat{\theta}_{\mathrm{ABS}}$ 和 $\hat{\theta}_{\mathrm{MAP}}$

8. (a) 考虑习题 7 中的观测量模型，其先验服从习题 5。计算基于 Y 的 Θ 的 MMSE 和 MMAE 估计。

(b) 计算(a)中的最小均方误差。

(c) 考虑如下观测模型

$$Y_k=N_k+\Theta,\quad k=1,\cdots,n$$

其中，N_1，N_2，…，N_n 和 Θ 是服从独立同分布的单位指数分布的随机变量。计算基于的 Θ 的 MAP。

9. 有一组观测量 Y_1，Y_2，…，Y_n 服从独立同分布，在 $\Theta=\theta$ 时服从给定的概率密度函数 p_θ，重做习题 1。

10. 推导式(4.2.47)。

11. 假设观测序列

$$Y_k=X_k+N_k,\quad k=1,\cdots,n$$

其中 N_1，N_2，…，N_n 为相互独立的均值为 0、方差为 σ^2 的高斯随机变量序列，X_1，X_2，…，X_n 由下式定义

$$X_0=\Theta$$

$$X_k=\alpha X_{k-1},\quad k=1,\cdots,n$$

其中，α 为已知的，Θ 为均值为 0、方差为 q^2 的高斯随机参数。

(a) 假设 Θ 和 N 是相互独立的，计算 Y_1，Y_2，…，Y_n 下 Θ 的 MMSE 估计。

(b) 对于每一个 $n=1, 2, \cdots$，令 $\hat{\theta}_n$ 表示在 $Y_1, Y_2, \cdots, Y_n$ 下 Θ 的 MMSE 估计。$\hat{\theta}_n$ 可以由下式计算

$$\hat{\theta}_n = K_n^{-1}[K_{n-1}\hat{\theta}_{n-1} + \alpha y_n], n = 1,2,\cdots$$

(c) 求取一个表达式，使得其均方误差为

$$e_n = E\{(\hat{\theta}_n - \Theta)^2\}, \quad n = 1,2,\cdots$$

当 $n\to\infty$、$q^2\to\infty$、$\sigma^2\to 0$、$\alpha<1$、$\alpha=1$、$\alpha>1$时，分别会有什么结论？

12. 假设 θ 是一个非随机参数且满足 $\theta>1$。另外，假设给定的 θ，$Y_1, Y_2, \cdots, Y_n$ 是独立同分布的观测量，其概率密度函数为

$$f_\theta(y)=\begin{cases}(\theta-1)y^{-\theta}, & y\geqslant 1\\ 0, & y\leqslant 1\end{cases}$$

求出 θ 的一个具有完备分布族的充分统计量，并对其进行证明。

13. 假设独立掷硬币 n 次，并定义观测序列为

$$Y_k = \begin{cases}1, & \text{如果第 } k \text{ 次结果是正面}\\ 0, & \text{如果第 } k \text{ 次结果是反面}\end{cases}$$

其中，$k=1, 2, \cdots, n$。令 $\theta=P(Y_k=1)$，$k=1, 2, \cdots, n$。

(a) 计算 θ 的 MVUE。

(b) 计算 θ 的 ML 估计，并计算其偏差和方差。

(c) 计算克拉美罗下限，并与(a)和(b)的结果作比较。

14. 推导式(4.4.22)。

15. 假设 Y 服从泊松分布，求出变化率的 ML 估计，计算偏差、方差和克拉美罗下限。

16. 假设 θ 是一个正的(非随机)参数，且有一组观测量 $Y_1, \cdots, Y_n$，其中，给定的 θ，$Y_1, \cdots, Y_n$ 服从独立同分布，其概率密度函数为

$$f_\theta(y)=\begin{cases}\dfrac{y^M e^{-y/2\theta}}{(2\theta)^{M+1}M!}, & y\geqslant 0\\ 0, & y<0\end{cases}$$

其中 M 是已知的正整数。

(a) 求出 θ 的 ML 估计。

(b) 计算(a)部分估计的偏差和方差。

(c) 计算 θ 无偏估计方差的克拉美罗下限。

(d) ML 估计是一致的吗？是否有效？

17. 假设观测两个联合高斯随机变量 Y_1 和 Y_2，均具有 0 均值和单位方差，需要估计相关系数$\rho=E\{Y_1Y_2\}$。

(a) 求出基于观测(Y_1, Y_2)的 ρ 的 ML 估计公式。

(b) 计算 ρ 的无偏估计的克拉美罗下限。

18. 假设存在一组观测序列 $Y_1, \cdots, Y_n$，满足

$$Y_k = N_k + \theta_{s_k}, k = 1,\cdots,n$$

其中，$\boldsymbol{N}=(N_1, \cdots, N_n)^{\mathrm{T}}$ 是 0 均值高斯随机向量，其协方差矩阵 $\boldsymbol{\Sigma}>0$，$s_1, \cdots, s_n$ 是一个已知信号序列，而且 θ 是(实的)非随机参数。

(a) 求出参数 θ 的 ML 估计。

(b) 计算(a)中结果的偏差和方差。

(c) 计算 θ 的无偏估计的克拉美罗下限，并与(b)的结果作比较。

(d) 当 $n\to\infty$ 时，$\hat{\theta}_{\mathrm{ML}}$ 的一致估计是什么？例如，假设有正常数 a 和 b 使得

$$\frac{1}{n}\sum_{k=1}^{n} s_k^2 > a, \quad \text{对于所有 } n \text{ 成立}$$

并且

$$\lambda_{\min}(\boldsymbol{\Sigma}^{-1})>b, \quad \text{对于所有 } n \text{ 成立}$$

其中，$\lambda_{\min}(\boldsymbol{\Sigma}^{-1})$表示矩阵 $\boldsymbol{\Sigma}^{-1}$ 的最小特征值。

19. 假设 θ 是一个正的非随机参数，有一组观测序列 $Y_1, \cdots, Y_n$ 满足

$$Y_k = \theta^{1/2}N_k, \quad k = 1,2,\cdots,n$$

其中，$\boldsymbol{N}=(N_1, \cdots, N_n)^{\mathrm{T}}$ 是 0 均值高斯随机向量，其协方差矩阵 $\boldsymbol{\Sigma}$。假设 $\boldsymbol{\Sigma}$ 是正定的。

(a) 找出基于 $Y_1, \cdots, Y_n$ 的 θ 的 ML 估计。

(b) 证明 ML 估计是无偏的。

(c) 计算 θ 的无偏估计的克拉美罗下限。

(d) 计算 θ 的 ML 估计的方差，并与克拉美罗下限作比较。

20. 考虑观测模型

$$Y_k = \theta^{1/2}s_kR_k + N_k, \quad k = 1,2,\cdots,n$$

其中，$s_1, s_2, \cdots, s_n$ 是已知信号，$N_1, \cdots, N_n, R_1, \cdots, R_n$ 是服从独立同分布的$\mathcal{N}(0, 1)$的随机变量，而且 $\theta\geqslant 0$ 是一个未知参数。

(a) 求出由 $Y_1, \cdots, Y_n$ 估计 θ 的似然公式。

(b) 求出 θ 的无偏估计方差的克拉美罗下限。

(c) 假设 $s_1, \cdots, s_n$ 是由 $+1$ 和 -1 组成的序列，求 θ 的 MLE。

(d) 计算(c)的偏差和方差，并将其方差值与克拉美罗下限作比较。

21. 假设 Y_1 和 Y_2 是包含参数 λ 的相互独立泊松随机变量，定义参数 θ 为

$$\theta = e^{-\lambda}$$

(a) 证明 (Y_1+Y_2) 是 θ 的一个完备充分统计量［假设 λ 的范围为 $(0, \infty)$］。

(b) 定义估计 $\hat{\theta}$ 为

$$\hat{\theta}(y) = \frac{1}{2}[f(y_1) + f(y_2)]$$

其中，定义 f 为

$$f(y) = \begin{cases} 1, & y=0 \\ 0, & y\neq 0 \end{cases}$$

证明 $\hat{\theta}$ 是 θ 的无偏估计。

(c) 求出 θ 的 MVUE（提示：(Y_1+Y_2) 是参数为 2λ 的泊松分布）。

(d) 求出 θ 的 MLE，判断该 MLE 值是否无偏并说明原因。

(e) 计算 θ 的无偏估计方差的克拉美罗下限。

22. 假设 $\theta>0$ 是感兴趣的参数，θ，Y_1，…，Y_n 是一组服从独立同分布的观测量且具有边缘分布函数

$$F_\theta(y) = [F(y)]^{1/\theta}, \quad -\infty < y < \infty$$

其中 F 是概率密度函数为 f 的已知分布函数。

(a) 证明下式是 θ 的 MVUE

$$\hat{\theta}_{\mathrm{MV}}(y) = -\frac{1}{n}\sum_{k=1}^{n}\log F(y_k)$$

(b) 利用下式的先验概率将 θ 替换为一个随机变量

$w(\theta) = c^m \exp(-c/\theta)/(\Gamma(m)\theta^{m+1}), \theta > 0$

其中，$c>0$ 和 $m>1$ 都是常数。利用 $E(\Theta)=c/(m-1)$，证明由 Y_1，…，Y_n 得到的 Θ 的 MMSE 估计是

$$\hat{\theta}_{\mathrm{MMSE}}(y) = \frac{c - \sum_{k=1}^{n}\log F(y_k)}{m+n-1}$$

(c) 根据先验信息比较 $\hat{\theta}_{\mathrm{MV}}$ 和 $\hat{\theta}_{\mathrm{MMSE}}$。

23. 假设已有观测

$$Y_k = A\sin\left(\frac{k\pi}{2} + \Phi\right) + N_k, \quad k = 1,\cdots,n$$

其中，$\mathbf{N}\sim\mathcal{N}(0, \sigma^2\mathbf{I})$ 是偶数。

(a) 假设 A 和 Φ 是非随机的，并且有 $A\geqslant 0$ 和 $\Phi\in[-\pi, \pi]$，求出对应的 ML 估计。

(b) 假设 A 和 Φ 是随机且独立的，其先验概率为

$$w_\Phi(\phi) = \begin{cases} \dfrac{1}{\pi}, & -\pi \leqslant \phi \leqslant \pi \\ 0, & \text{其他} \end{cases}$$

$$w_A(a) = \begin{cases} (a/\beta^2)e^{-a^2/2\beta^2}, & a \geqslant 0 \\ 0, & a < 0 \end{cases}$$

其中 β 为已知参数。假设 A 和 Φ 独立于 $\mathbf{N}$，找出 A 和 Φ 的 MAP 估计。

(c) 在什么条件下(a)和(b)的估计近似相等？

24. 假设给定 $\Theta=\theta$，Y_1，…，Y_n 是服从独立同分布的实观测，其边缘密度函数为

$$f_\theta(y) = \begin{cases} \theta^{-1}e^{-y/\theta}, & y\geqslant 0 \\ 0, & y<0 \end{cases}$$

(a) 求出基于 Y_1，…，Y_n 的 θ 的 ML 估计，计算其均值和方差。

(b) 计算 θ 的无偏估计方差的克拉美罗下限。

(c) 假设 Θ 均匀分布在 $(0, 1]$ 范围，求出 Θ 的 MAP 估计。

(d) $n=3$ 时，求出 Θ 的 MMSE 估计，假设具有与(c)相同的先验概率。

(e) $n=2$ 时，求出 Θ 的 MMSE 估计，假设具有与(c)相同的先验概率。

25. 假设对于给定的 $\Theta=\theta$，实观测 Y 具有以下概率密度函数

$$p_\theta(y) = \begin{cases} \dfrac{6(y^2+\theta y)}{2+3\theta}, & 0 \leqslant y \leqslant 1 \\ 0, & \text{其他} \end{cases}$$

(a) 假设 Θ 均匀分布在 $[0, 1]$，求出 MMSE 估计及其对应的最小贝叶斯风险。

(b) 对于(a)中的 Θ，求出 Θ 的 MAP 估计和 MMSE 估计。

(c) 求出 θ 的最大似然估计并计算其偏差。

(d) 计算 θ 的无偏估计方差的克拉美罗下限。

第5章 信号估计基础

5.1 引言

第 4 章讨论了静态参数估计器的设计方法，即时不变参数的估计器设计方法。而在许多实际应用中，对于动态或者时变参数的估计问题也很感兴趣。在传统术语中，动态参数也称为信号，因此对于时变参数的估计问题也称为信号估计或者跟踪问题。

许多实际应用中都会出现信号估计问题。例如，雷达系统的功能之一是对通过雷达扫描区域的目标进行跟踪。这意味着雷达必须在连续的时间内估计目标的位置(也可能是速度)。由于目标通常是运动的，且位置测量受噪声影响，因此这是典型的信号估计问题。模拟通信是信号估计的另一典型应用，它一般采用调幅、调频或者调相的正弦波来传输模拟信息(包括音频、视频等)。在这种情况下，接收机的作用是从包含噪声的接收波形中尽可能高保真地检测出发射信号。同理，由于发送信号具有时变性，这也属于信号估计的问题。

在信号估计中，对检测模型进行统计建模时需要考虑参数的动态特性，尤其是需要从统计意义上对信号的动态特性(变化快慢，以哪种方式变化等)进行建模，以获得有意义的信号估计过程。动态参数估计器的预期性能也不同于静态参数。特别地，由于信号的时变性，不能期望通过观测无穷多样本以得到信号的准确估计。

本章讨论信号估计的基本思想。5.2 节讨论 Kalman-Bucy 滤波，它是通过有限维线性动态模型综合得到的，是一种非常实用的信号估计算法。5.3 节介绍对观测样本进行线性变换的信号估计方法。5.4 节介绍线性估计的一个特例，即 Wiener-Kolmogorov 滤波器，它是估计时域平稳信号的一种方法。

5.2 Kalman-Bucy 滤波

随时间变化的物理量可建模成如下形式

$$\boldsymbol{X}_{n+1} = \boldsymbol{f}_n(\boldsymbol{X}_n, \boldsymbol{U}_n), \quad n = 0, 1, \cdots \tag{5.2.1}$$

其中，$\boldsymbol{X}_0$，$\boldsymbol{X}_1$，$\cdots \in \mathbf{R}^m$ 代表研究对象，它们是 m 维实空间的向量序列；$\boldsymbol{U}_0$，$\boldsymbol{U}_1$，$\cdots \in \mathbf{R}^s$ 是 s 维实空间的激励向量序列；$\boldsymbol{f}_0$，$\boldsymbol{f}_1$，$\cdots$表示将 $\mathbf{R}^m \times \mathbf{R}^s$ 映射到 $\mathbf{R}^m$ 的函数序列(可以是时变函数)。式(5.2.1)是动态系统的典型例子，每个 $\boldsymbol{X}_n$ 代表的是系统在时刻 n 的状态，并且 $\boldsymbol{U}_n$ 代表的是在时刻 n 对系统的输入[参见 Desoer(1970)]。在一个动态系统中，对于任意固定的时刻 k 和 l，$\boldsymbol{X}_l$ 将被 k 时刻的状态($\boldsymbol{X}_k$)和 k 时刻至 $l-1$ 时刻的系统输入($\{\boldsymbol{U}_n\}_{n=k}^{l-1}$)唯一确定。注意，利用式(5.2.1)估计$\{\boldsymbol{X}_n\}_{n=1}^{\infty}$不仅需要一定形式的输入激励序列，而且需要初始条件 $\boldsymbol{X}_0$ 也具有某种形式。如果输入或者初始条件是随机的，那么系统的状态向量 $\boldsymbol{X}_0$，$\boldsymbol{X}_1$，$\cdots$为随机向量，此时式(5.2.1)为随机系统。

式(5.2.1)描述了一个系统的状态更新过程，也称为系统的状态方程。系统通常还有一个 k 维输出序列 $\boldsymbol{Z}_0$，$\boldsymbol{Z}_1$，$\cdots \in \mathbf{R}^k$，输出序列由输出方程给出

$$\boldsymbol{Z}_n = \boldsymbol{h}_n(\boldsymbol{X}_n), \quad n = 0,1,\cdots \tag{5.2.2}$$

其中，$\boldsymbol{h}_n$ 将 $\mathbf{R}^m$ 映射到 $\mathbf{R}^k$。因此整个系统完成从初始状态 $\boldsymbol{X}_0$ 和输入序列$\{\boldsymbol{U}_n\}_{n=n}^{\infty}$到输出序列$\{\boldsymbol{Z}_n\}_{n=0}^{\infty}$的映射。

下面将给出由式(5.2.1)和式(5.2.2)所描述的系统的例子。

例 5.2.1 一维运动

假设需要对一个在 $t>0$ 时具有加速度 A_t 的质点的运动过程进行建模。在任意时刻 t，质点的位置坐标 P_t 和速度 V_t 分别满足方程 $V_t = \mathrm{d}P_t/\mathrm{d}t$ 和 $A_t = \mathrm{d}V_t/\mathrm{d}t$。假设每间隔 T_s 时间观察一次质点的位置坐标，并为其建立形如式(5.2.1)和(5.2.2)的模型，以此来观察质点从一个时刻到另一个时刻的位置变化。假设观察时间间隔 T_s 非常短，其一阶泰勒展开形式为

$$P_{(n+1)T_s} \approx P_{nT_s} + T_s V_{nT_s} \tag{5.2.3a}$$

$$V_{(n+1)T_s} \approx V_{nT_s} + T_s A_{nT_s} \tag{5.2.3b}$$

从式(5.2.3)可以看出，需要两个状态变量来描述质点的运动，即位置和速度。定义 $Z_n = X_{1,n} = P_{nT_s}$，$X_{2,n} = V_{nT_s}$，$U_n = A_{nT_s}$，那么一维运动可以用状态方程近似表示为

$$\boldsymbol{X}_{n+1} = \boldsymbol{F}\boldsymbol{X}_n + \boldsymbol{G}\boldsymbol{U}_n, \quad n = 0,1,\cdots \tag{5.2.4}$$

输出方程为

$$Z_n = \boldsymbol{H}\boldsymbol{X}_n, \quad n = 0,1,\cdots \tag{5.2.5}$$

其中，$\boldsymbol{F}$ 是一个 2×2 的矩阵

$$\boldsymbol{F} = \begin{pmatrix} 1 & T_s \\ 0 & 1 \end{pmatrix} \tag{5.2.6}$$

$\boldsymbol{G}$ 是 2×1 的矩阵

$$\boldsymbol{G} = \begin{pmatrix} 0 \\ T_s \end{pmatrix} \tag{5.2.7}$$

$\boldsymbol{H}$ 是 1×2 的矩阵

$$\boldsymbol{H} = (1 \quad 0) \tag{5.2.8}$$

因此，本例中有 $m=2$，$s=1$，$k=1$，并且函数 $\boldsymbol{f}_n$ 和 $\boldsymbol{h}_n$ 分别为

$$\boldsymbol{f}_n(\boldsymbol{X},\boldsymbol{U}) == \boldsymbol{F}\boldsymbol{X} + \boldsymbol{G}\boldsymbol{U} \tag{5.2.9}$$

和

$$\boldsymbol{h}_n(\boldsymbol{X}) = \boldsymbol{H}\boldsymbol{X} \tag{5.2.10}$$

下面将进一步讨论这个特殊模型。◀

在许多实际应用中，经常会面对如下问题：随机系统一段时间(t 时间内)的输出通常伴有观测噪声(或测量噪声)，需要估计系统 u 时刻的状态。对应的观测序列为

$$\boldsymbol{Y}_n = \boldsymbol{Z}_n + \boldsymbol{V}_n, \quad n = 0,1,\cdots,t \tag{5.2.11}$$

需要通过这个序列估计 $\boldsymbol{X}_u$。在式(5.2.11)中，序列 $\boldsymbol{V}_0$，$\boldsymbol{V}_1$，$\cdots$ 表示测量噪声。式(5.2.11)通常称为测量方程。若 $u=t$，那么这个估计问题为滤波问题；若 $u<t$，称为平滑问题；若 $u>t$，则为预测问题。对于这三类问题，均可采用状态估计这一术语。

如前所述，状态估计问题在许多实际应用中都会出现。例如，在“边跟踪边扫描”雷达中，雷达对目标位置的测量是在每次扫描时刻获得的。与例 5.2.1(具有随机加速度)相似，这种测量实际上也是在随机系统噪声条件下完成的。在每次扫描时刻，雷达希望能够对目标当前时刻的位置进行估计，并且预测目标下一个时刻的位置。即在扫描时刻 t，根据目标过去时刻的状态，估计目标在 $u=t$ 和 $u=t+1$ 时刻的状态(后面将会进一步讨论这种特殊的应用)。

状态估计在自动控制系统中也有所应用，如飞机飞行控制、化学过程控制等。在飞行控制中，状态是飞机的位置坐标和描述飞机角度偏转的姿态坐标(即横滚、俯仰和偏航)。在这种情况下，状态方程描述的是飞机的动态过程，此时的输入可能是施加在飞机上的控制力和随机力(例如湍流)。在化学过程控制中，状态变量可能是温度和其他化学物品的浓度，状态方程描述的是一个化学反应过程。当然，许多其他应用实例也可以应用本章的模型进行描述。

如果对上述模型中的状态估计 $\hat{\boldsymbol{X}}_u$ 采用均方误差性能 $E\{\|\hat{\boldsymbol{X}}_u-\boldsymbol{X}_u\|^2\}$ 来衡量，从第 4 章(模型 4.2.4)可知，最优估计为条件均值

$$\hat{\boldsymbol{X}}_u = E\{\boldsymbol{X}_u \mid \boldsymbol{Y}_0,\cdots,\boldsymbol{Y}_t\} \tag{5.2.12}$$

当然，对于固定的时刻 u 和 t，这个问题和第 4 章讨论的向量估计问题相同。随着时间的增加，如何对信号进行实时估计就显得尤为重要。数据随时间 t 线性增加，式(5.2.12)得到的条件均值估计将失去实际意义，除非式(5.2.12)的估计器具有特殊结构，使得信号估计过程能够高效完成。因此，在进一步考虑式(5.2.12)之前，需要对式(5.2.1)、式(5.2.2)、式(5.2.11)的模型进行一些合适的限制约束。

一种约束是假设系统为一个线性随机系统，此时状态方程和观测方程的形式为

$$\boldsymbol{X}_{n+1} = \boldsymbol{F}_n\boldsymbol{X}_n + \boldsymbol{G}_n\boldsymbol{U}_n, \quad n = 0,1,\cdots \tag{5.2.13a}$$

$$\boldsymbol{Y}_n = \boldsymbol{H}_n\boldsymbol{X}_n + \boldsymbol{V}_n, \quad n = 0,1,\cdots \tag{5.2.13b}$$

其中，$\boldsymbol{F}_n$、$\boldsymbol{G}_n$ 和 $\boldsymbol{H}_n$ 的维数分别是 $m\times m$、$m\times s$ 和 $k\times m$。式(5.2.13a)的线性模型对于很多应用都是适用的。例如，在例 5.2.1 的一维运动模型(二维和三维与之类似)中，系统是一个加速度为随机变量的线性随机系统。对于许多非线性系统而言，当状态变量表示系统实际轨迹与期望轨迹的偏差时，那么这个非线性系统可以采用线性系统进行逼近。特别地，通过对非线性函数 $\boldsymbol{f}_n$ 进行泰勒级数展开，可将许多系统线性化。

为简化式(5.2.12)，进一步假设输入序列 $\{\boldsymbol{U}_n\}_{n=0}^{\infty}$ 和观测噪声序列 $\{\boldsymbol{V}_n\}_{n=0}^{\infty}$ 是独立的零均值高斯随机向量。此外，假设初始条件 $\boldsymbol{X}_0$ 也是高斯随机向量，且独立于 $\{\boldsymbol{U}_n\}_{n=0}^{\infty}$ 和 $\{\boldsymbol{V}_n\}_{n=0}^{\infty}$。事实上，序列 $\{\boldsymbol{U}_n\}_{n=0}^{\infty}$ 和 $\{\boldsymbol{V}_n\}_{n=0}^{\infty}$ 独立的假设条件可以适当放宽，零均值的假设也仅仅是为了简化分析。相反，高斯假设是非常关键的。并且，在许多模型中，这些假设都是合理的。因为观测噪声经常是由电子传感器中的高斯热噪声引起，而由湍流引起的随机输入通常也符合高斯统计特性。此外，事实证明，如果可以接受所有线性估计器中的最优估计器，那么高斯假设也可以适当放宽。下面将具体讨论这个问题。

基于上述假设，从计算效率的角度来看，式(5.2.12)中的条件均值状态估计器具有很好的形式。这个形式在其他的许多状态估计问题中已经出现，本节考虑滤波问题($u=t$)和

单步预测问题($u=t+1$)，因为滤波和单步预测在实际应用中最常见。下文给出了这两个问题的具体描述。

命题 5.2.1 离散时间 Kalman-Bucy 滤波

在式(5.2.13)所描述的线性随机系统中，假设$\{\boldsymbol{U}_n\}_{n=0}^{\infty}$和$\{\boldsymbol{V}_n\}_{n=0}^{\infty}$是相互独立的零均值高斯向量，且独立于高斯初始状态$\boldsymbol{X}_0$，则估计器$\hat{\boldsymbol{X}}_{t|t}\triangleq E\{\boldsymbol{X}_t|\boldsymbol{Y}_0^t\}$[⊖]和$\hat{\boldsymbol{X}}_{t+1|t}\triangleq E\{\boldsymbol{X}_{t+1}|\boldsymbol{Y}_0^t\}$的递归计算式为

$$\hat{\boldsymbol{X}}_{t|t}=\hat{\boldsymbol{X}}_{t|t-1}+\boldsymbol{K}_t(\boldsymbol{Y}_t-\boldsymbol{H}_t\hat{\boldsymbol{X}}_{t|t-1}),\quad t=0,1,\cdots \tag{5.2.14a}$$

和

$$\hat{\boldsymbol{X}}_{t+1|t}=\boldsymbol{F}_t\hat{\boldsymbol{X}}_{t|t},\quad t=0,1,\cdots \tag{5.2.14b}$$

其中，初始条件$\hat{\boldsymbol{X}}_{0|-1}=\boldsymbol{m}_0\triangleq E\{X_0\}$。$\boldsymbol{\Sigma}_{t|t-1}\triangleq\mathrm{cov}(\boldsymbol{X}_t|\boldsymbol{Y}_0^{t-1})$，$\boldsymbol{R}_t\triangleq\mathrm{cov}(\boldsymbol{V}_t)$，矩阵$\boldsymbol{K}_t$可表示为

$$\boldsymbol{K}_t=\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}_t^T(\boldsymbol{H}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}_t^T+\boldsymbol{R}_t)^{-1} \tag{5.2.15}$$

注意，$\hat{\boldsymbol{X}}_{t|t-1}=E\{\boldsymbol{X}_t|\boldsymbol{Y}_0^{t-1}\}$，即$\boldsymbol{\Sigma}_{t|t-1}$是$\boldsymbol{Y}_0^{t-1}$条件下预测误差$\boldsymbol{X}_t-\hat{\boldsymbol{X}}_{t|t-1}$的协方差矩阵。该矩阵和滤波误差矩阵$\boldsymbol{\Sigma}_{t|t}\triangleq\mathrm{cov}(\boldsymbol{X}_t|\boldsymbol{Y}_0^t)$的递归计算式为

$$\boldsymbol{\Sigma}_{t|t}=\boldsymbol{\Sigma}_{t|t-1}-\boldsymbol{K}_t\boldsymbol{H}_t\boldsymbol{\Sigma}_{t|t-1},\quad t=0,1,\cdots \tag{5.2.16a}$$

$$\boldsymbol{\Sigma}_{t+1|t}=\boldsymbol{F}_t\boldsymbol{\Sigma}_{t|t}\boldsymbol{F}_t^T+\boldsymbol{G}_t\boldsymbol{Q}_t\boldsymbol{G}_t^T,\quad t=0,1,\cdots \tag{5.2.16b}$$

其中，初始矩阵为$\boldsymbol{\Sigma}_{0|-1}=\boldsymbol{\Sigma}_0=\mathrm{cov}(\boldsymbol{X}_0)$，$\boldsymbol{Q}_t$是$t$时刻状态输入的协方差矩阵，即$\boldsymbol{Q}_t=\mathrm{cov}(\boldsymbol{U}_t)$。

证明 为了证明这个命题的正确性，首先直接给出式(5.2.14b)和式(5.2.16b)，然后利用归纳法证明式(5.2.14a)和式(5.2.16a)。结合状态方程，式(5.2.14b)可表示为

$$\begin{aligned}\hat{\boldsymbol{X}}_{t+1|t}&=E\{\boldsymbol{X}_{t+1}|\boldsymbol{Y}_0^t\}=E\{\boldsymbol{F}_t\boldsymbol{X}_t+\boldsymbol{G}_t\boldsymbol{U}_t|\boldsymbol{Y}_0^t\}\\&=\boldsymbol{F}_tE\{\boldsymbol{X}_t|\boldsymbol{Y}_0^t\}+\boldsymbol{G}_tE\{\boldsymbol{U}_t|\boldsymbol{Y}_0^t\}\\&=\boldsymbol{F}_t\hat{\boldsymbol{X}}_{t|t}+\boldsymbol{G}_tE\{\boldsymbol{U}_t|\boldsymbol{Y}_0^t\}\end{aligned} \tag{5.2.17}$$

其中，第三个等式根据期望的线性性质得到，最后一个等式由$\hat{\boldsymbol{X}}_{t|t}$的定义直接得到。注意，$\boldsymbol{Y}_0^t$由$\boldsymbol{X}_0^t$和$\boldsymbol{V}_0^t$确定，或者由$\boldsymbol{X}_0$、$\boldsymbol{U}_0^{t-1}$和$\boldsymbol{V}_0^t$确定，它们都与$\boldsymbol{U}_t$相独立。换言之，式(5.2.17)中的第二项是条件无关的，即$E\{\boldsymbol{U}_t|\boldsymbol{Y}_0^t\}=E\{\boldsymbol{U}_t\}=0$。代入式(5.2.17)后有式(5.2.14b)成立。类似地，考虑$\boldsymbol{U}_t$独立于$\boldsymbol{X}$和$\boldsymbol{Y}_0^t$，可得

$$\begin{aligned}\boldsymbol{\Sigma}_{t+1|t}&=\mathrm{cov}(\boldsymbol{X}_{t+1}|\boldsymbol{Y}_0^t)=\mathrm{cov}(\boldsymbol{F}_t\boldsymbol{X}_t+\boldsymbol{G}_t\boldsymbol{U}_t|\boldsymbol{Y}_0^t)\\&=\mathrm{cov}(\boldsymbol{F}_t\boldsymbol{X}_t|\boldsymbol{Y}_0^t)+\mathrm{cov}(\boldsymbol{G}_t\boldsymbol{U}_t|\boldsymbol{Y}_0^t)\\&=\mathrm{cov}(\boldsymbol{F}_t\boldsymbol{X}_t|\boldsymbol{Y}_0^t)+\mathrm{cov}(\boldsymbol{G}_t\boldsymbol{U}_t)\end{aligned} \tag{5.2.18}$$

利用$\mathrm{cov}(\boldsymbol{AX})=\boldsymbol{A}\mathrm{cov}(\boldsymbol{X})A^{\mathrm{T}}$并结合$\boldsymbol{\Sigma}_{t|t}$和$\boldsymbol{Q}_t$的定义，可得

$$\boldsymbol{\Sigma}_{t+1|t}=\boldsymbol{F}_t\mathrm{cov}(\boldsymbol{X}_t|\boldsymbol{Y}_0^t)\boldsymbol{F}_t^T+\boldsymbol{G}_t\mathrm{cov}(\boldsymbol{U}_t)\boldsymbol{G}_t^T=\boldsymbol{F}_t\boldsymbol{\Sigma}_{t|t}\boldsymbol{F}_t^T+\boldsymbol{G}_t\boldsymbol{Q}_t\boldsymbol{G}_t^T \tag{5.2.19}$$

即式(5.2.16b)得证。

至此已经证明式(5.2.14b)和式(5.2.16b)成立。现在采用归纳法证明式(5.2.14a)和式(5.2.16a)也成立。为此，必须证明对于$t=0$和任意的$t_0>0$，当$t=t_0-1$成立时$t=t_0$

⊖ 为便于表示，将集合$\boldsymbol{Y}_a$，…，$\boldsymbol{Y}_b$表示为$\boldsymbol{Y}_a^b(b>a)$。

也成立。当$t=0$时，测量方程为

$$\boldsymbol{Y}_0=\boldsymbol{H}_0\boldsymbol{X}_0+\boldsymbol{V}_0 \tag{5.2.20}$$

由于$\boldsymbol{X}_0$和$\boldsymbol{V}_0$是相互独立的高斯向量，利用$\boldsymbol{Y}_0$对$\boldsymbol{X}_0$进行估计符合例4.2.3中讨论的线性估计模型。特别地，由于$\boldsymbol{X}_0\sim\mathcal{N}(\boldsymbol{m}_0,\boldsymbol{\Sigma}_0)$和$\boldsymbol{V}_0\sim\mathcal{N}(0,\boldsymbol{R}_0)$，由式(4.2.53)可以得到

$$\begin{aligned}\hat{\boldsymbol{X}}_{0\mid 0}&\triangleq E\{\boldsymbol{X}_0\mid\boldsymbol{Y}_0\}=\boldsymbol{m}_0+\boldsymbol{\Sigma}_0\boldsymbol{H}_0^T(\boldsymbol{H}_0\boldsymbol{\Sigma}_0\boldsymbol{H}_0^T+\boldsymbol{R}_0)^{-1}(\boldsymbol{Y}_0-\boldsymbol{H}_0\boldsymbol{m}_0)\\&=\hat{\boldsymbol{X}}_{0\mid -1}+\boldsymbol{K}_0(\boldsymbol{Y}_0-\boldsymbol{H}_0\hat{\boldsymbol{X}}_{0\mid -1})\end{aligned} \tag{5.2.21}$$

其中，$\hat{\boldsymbol{X}}_{0\mid -1}=\boldsymbol{m}_0$，$\boldsymbol{K}_0=\boldsymbol{\Sigma}_{0\mid -1}\boldsymbol{H}_0^T(\boldsymbol{H}_0\boldsymbol{\Sigma}_{0\mid -1}\boldsymbol{H}_0^T+\boldsymbol{R}_0)^{-1}$和$\boldsymbol{\Sigma}_{0\mid -1}=\boldsymbol{\Sigma}_0$与定理中的定义相同。式(5.2.21)是式(5.2.14a)在$t=0$时的情形。从式(5.2.21)得到的协方差矩阵可表示为

$$\boldsymbol{\Sigma}_{0\mid 0}=\boldsymbol{\Sigma}_0-\boldsymbol{\Sigma}_0\boldsymbol{H}_0^T(\boldsymbol{H}_0\boldsymbol{\Sigma}_0\boldsymbol{H}_0^T+\boldsymbol{R}_0)^{-1}\boldsymbol{H}_0\boldsymbol{\Sigma}_0=\boldsymbol{\Sigma}_{0\mid -1}-\boldsymbol{K}_0\boldsymbol{H}_0\boldsymbol{\Sigma}_{0\mid -1} \tag{5.2.22}$$

式(5.2.22)对应式(5.2.16a)在$t=0$时的形式。

为完成证明过程，假设式(5.2.14a)和式(5.2.16a)在$t=t_0-1$时成立。由于$\boldsymbol{X}_{t_0}$和$\boldsymbol{Y}_0^{t_0-1}$是由高斯向量$\boldsymbol{X}_0$、$\boldsymbol{U}_0^{t_0-1}$和$\boldsymbol{V}_0^{t_0-1}$经过线性变换得到的，因此$\boldsymbol{X}_{t_0}$和$\boldsymbol{Y}_0^{t_0-1}$服从联合高斯分布，且给定$\boldsymbol{Y}_0^{t_0-1}$，$\boldsymbol{X}_{t_0}$服从条件高斯分布。特别地，在给定$\boldsymbol{Y}_0^{t_0-1}$的条件下，$\boldsymbol{X}_{t_0}$的条件分布为$\boldsymbol{X}_{t_0}\sim\mathcal{N}(\hat{\boldsymbol{X}}_{t_0\mid t_0-1},\boldsymbol{\Sigma}_{t_0\mid t_0-1})$。此外，$\boldsymbol{V}_{t_0-1}$也是独立于$\boldsymbol{Y}_0^{t_0-1}$的高斯变量，因此在给定$\boldsymbol{Y}_0^{t_0-1}$的情况下，也可认为它是服从条件高斯分布的，且其条件分布为$\mathcal{N}(\boldsymbol{0},\boldsymbol{R}_{t_0})$。又因为$\boldsymbol{V}_{t_0}$独立于$\boldsymbol{X}_0$、$\boldsymbol{U}_0^{t_0-1}$和$\boldsymbol{V}_0^{t_0-1}$，因此在给定$\boldsymbol{Y}_0^{t_0-1}$的情况下，它也是条件独立于$\boldsymbol{X}_{t_0}$的。综上所述，在给定$\boldsymbol{Y}_0^{t_0-1}$的条件下，观测方程为

$$\boldsymbol{Y}_{t_0}=\boldsymbol{H}_{t_0}\boldsymbol{X}_{t_0}+\boldsymbol{V}_{t_0} \tag{5.2.23}$$

式(5.2.23)为高斯线性方程，与例4.2.3中所讨论的方程具有相同的形式。在条件模型式(5.2.23)中，通过计算给定$\boldsymbol{Y}_{t_0}$情况下$\boldsymbol{X}_{t_0}$的条件期望，可得$\hat{\boldsymbol{X}}_{t_0\mid t_0}$。利用式(4.2.53)，有

$$\hat{\boldsymbol{X}}_{t_0\mid t_0}=\hat{\boldsymbol{X}}_{t_0\mid t_0-1}+\Sigma_{t_0\mid t_0-1}\boldsymbol{H}_{t_0}^T(\boldsymbol{H}_{t_0}\Sigma_{t_0\mid t_0-1}\boldsymbol{H}_{t_0}^T+\boldsymbol{R}_{t_0})^{-1}(\boldsymbol{Y}_{t_0}-\boldsymbol{H}_{t_0}\hat{\boldsymbol{X}}_{t_0\mid t_0-1}) \tag{5.2.24}$$

这里应用了一个事实，即在给定$\boldsymbol{Y}_0^{t_0-1}$的条件下，$\hat{\boldsymbol{X}}_{t_0}$的条件分布是$N(\hat{\boldsymbol{X}}_{t_0\mid t_0-1},\boldsymbol{\Sigma}_{t_0\mid t_0-1})$。结合$\boldsymbol{K}_{t_0}$的定义易知：式(5.2.24)是式(5.2.14a)在$t=t_0$时的特殊情形。类似地，通过式(4.2.54)和上面的结论，可得式(5.2.16a)成立。由此$\hat{\boldsymbol{X}}_{t_0\mid t_0}$可由式(5.2.14a)利用$\hat{\boldsymbol{X}}_{t_0\mid t_0-1}$得到，公式$\boldsymbol{\Sigma}_{t_0\mid t_0}$可以通过式(5.2.16a)利用$\boldsymbol{\Sigma}_{t_0\mid t_0-1}$得到。并且假设式(5.2.14a)和式(5.2.16a)对于$t=t_0-1$是成立的，可得$t=t_0$也成立。至此完成命题的全部证明。 □

命题5.2.1所描述的估计方法称为离散时间Kalman-Bucy滤波器，因为它是R. E. Kalman和R. S. Bucy在20世纪50年代末提出的连续时间递归状态估计的离散化表示。图5.2.1中给出了这一估计方法的结构，可以看出这种结构的计算非常简便。特别地，虽然估计器$\hat{\boldsymbol{X}}_{t+1\mid t}$或者$\hat{\boldsymbol{X}}_{t\mid t}$依赖于全部数据$\boldsymbol{Y}_0^t$，但是在每一步中它们只由最近的观测值$\boldsymbol{Y}_t$和前一个预测值$\hat{\boldsymbol{X}}_{t\mid t-1}$决定。因此，与存储$(t+1)k$维向量$\boldsymbol{Y}_0$(这将带来线性增大的存储需要和计算负担)不同，该估计器仅需要存储和更新m维向量$\hat{\boldsymbol{X}}_{t\mid t-1}$。估计器的其他部分(包括卡尔曼增益矩阵$\boldsymbol{K}_t$)独立于观测数据，仅与模型参数有关。

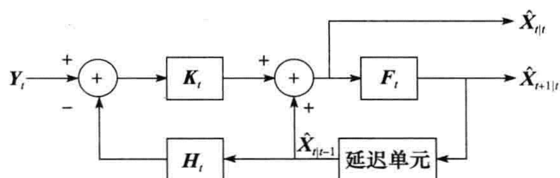


图5.2.1 离散时间Kalman-Bucy滤波器

递归公式(5.2.14)和式(5.2.16)都包含两个基本步骤。第一步是**测量值更新**[式(5.2.14a)和式(5.2.16a)]，它们提供了在给定$\boldsymbol{Y}_0^{t-1}$和当前观测值$\boldsymbol{Y}_t$的条件下，$\hat{\boldsymbol{X}}_t$的状态估计和协方差矩阵估计的更新方法。第二步为**时间更新**[式(5.2.14b)和式(5.2.16b)]，它提供了根据当前时刻测量值$\boldsymbol{Y}_0^t$，将估计的状态和协方差矩阵投影到下一时刻$t+1$的更新方法。考察式(5.2.1)中定理的证明过程，会发现时间更新完全来自状态方程，而测量值更新来自测量方程。

进一步考虑测量更新方程(5.2.14a)。特别地，估计值$\hat{\boldsymbol{X}}_{t|t}$是在给定$\boldsymbol{Y}_0^t$的条件下$\boldsymbol{X}_t$的最优估计器，也可以看作是基于过去时间$\hat{\boldsymbol{X}}_{t|t-1}$得到的$\boldsymbol{X}_t$的最优估计器与一个修正项$\boldsymbol{K}_t(\boldsymbol{Y}_t-\boldsymbol{H}_t\hat{\boldsymbol{X}}_{t|t-1})$的组合。在修正项中出现的向量$\boldsymbol{I}_t\triangleq\boldsymbol{K}_t(\boldsymbol{Y}_t-\boldsymbol{H}_t\hat{\boldsymbol{X}}_{t|t-1})$有一个非常有意思的解释。已知$\boldsymbol{Y}_t=\boldsymbol{H}_t\boldsymbol{X}_t+\boldsymbol{V}_t$，且$\boldsymbol{Y}_{t|t-1}\triangleq E\{\boldsymbol{Y}_t\mid\boldsymbol{Y}_0^{t-1}\}+\boldsymbol{H}_tE\{\boldsymbol{X}_t\mid\boldsymbol{Y}_0^{t-1}\}+E\{\boldsymbol{V}_t\mid\boldsymbol{Y}_0^{t-1}\}=\boldsymbol{H}_t\hat{\boldsymbol{X}}_{t|t-1}$，这里应用了$\boldsymbol{V}_t$独立于$\boldsymbol{Y}_0^{t-1}$且均值为0的性质。因此$\boldsymbol{I}_t=\boldsymbol{Y}_t-\hat{\boldsymbol{Y}}_{t|t-1}$表示由过去值$\boldsymbol{Y}_0^{t-1}$预测$\boldsymbol{Y}_t$时产生的误差。该误差通常称为**预测残余**或者**新息**，后一种叫法来源于这一事实：将$\boldsymbol{Y}_t$写为

$$\boldsymbol{Y}_t=\hat{\boldsymbol{Y}}_{t|t-1}+\boldsymbol{I}_t \tag{5.2.25}$$

$\hat{\boldsymbol{Y}}_{t|t-1}$可理解为$\boldsymbol{Y}_t$中可以由过去的量估计得到的部分，而$\boldsymbol{I}_t$代表$\boldsymbol{Y}_t$中不能被预测的部分。这样，$\boldsymbol{I}_t$中包含的是由第$t$个观测值带来的新的信息，因此称为“新息”。(这个序列在第3章的高斯检测问题中也出现过。)

不难发现，新息序列$\{\boldsymbol{I}_t\}_{t=0}^{\infty}$是一个零均值的独立高斯随机向量。首先，$\{\boldsymbol{I}_t\}_{t=0}^{\infty}$是高斯序列，这是因为$\{\boldsymbol{Y}_t\}_{t=0}^{\infty}$是高斯序列并且$\{\boldsymbol{I}_t\}_{t=0}^{\infty}$是$\{\boldsymbol{Y}_t\}_{t=0}^{\infty}$的线性变换。其次，应用条件均值的均值迭代性质($E\{Y\}=E\{E\{Y\mid X\}\}$)，可得$\boldsymbol{I}_t$的均值为

$$E\{\boldsymbol{I}_t\}=E\{\boldsymbol{Y}_t-E\{\boldsymbol{Y}_t\mid\boldsymbol{Y}_0^{t-1}\}\}=E\{\boldsymbol{Y}_t\}-E\{\boldsymbol{Y}_t\}=0$$

最后，利用$E\{\boldsymbol{I}_t\}=0$，有

$$\mathrm{cov}(\boldsymbol{I}_t,\boldsymbol{I}_s)=E\{\boldsymbol{I}_t\boldsymbol{I}_s^T\}$$

假设$s<t$，得

$$E\{\boldsymbol{I}_t\boldsymbol{I}_s^T\}=E\{E\{\boldsymbol{I}_t\boldsymbol{I}_s^T\mid\boldsymbol{Y}_0^s\}\}=E\{E\{\boldsymbol{I}_t\mid\boldsymbol{Y}_0^s\}\boldsymbol{I}_s^T\} \tag{5.2.26}$$

其中，第二个等式的根据为$\boldsymbol{I}_s$在给定$\boldsymbol{Y}_0^s$的条件下是一个常数。注意

$$E\{\boldsymbol{I}_t|\boldsymbol{Y}_0^s\}=E\{\boldsymbol{Y}_t|\boldsymbol{Y}_0^s\}-E\{E\{\boldsymbol{Y}_t|\boldsymbol{Y}_0^{t-1}\}\boldsymbol{Y}_0^s\}=E\{\boldsymbol{Y}_t|\boldsymbol{Y}_0^s\}-E\{\boldsymbol{Y}_t|\boldsymbol{Y}_0^s\}=\boldsymbol{0}$$

可见$\mathrm{cov}(\boldsymbol{I}_t,\boldsymbol{I}_s)=0$。对于$s>t$，根据对称参数可得相同的结果。因此，新息向量间是无关的，又因为它们服从联合高斯分布，因此它们是独立的。

综合以上讨论以及式(5.2.25)，可以再次明确：$\boldsymbol{Y}_t$包含两部分，$\hat{\boldsymbol{Y}}_{t|t-1}$是完全依赖于过去时刻的可预测部分，而另一部分$\boldsymbol{I}_t$完全独立于过去时刻。这意味着新息序列提供了一组独立的观测序列，且这一观测序列等价于原始集合$\{\boldsymbol{Y}_t\}_{t=0}^{\infty}$。换言之，构造新息序列对应估计器的预白化过程，这在第3章中也提到过[⊖]。

下面举例说明 Kalman-Bucy 的特性。

⊖ 注意，向量$\boldsymbol{I}_t$并非独立分布的。易知$\mathrm{cov}(\boldsymbol{I}_t)=\boldsymbol{H}_t\boldsymbol{\Sigma}_{t/t-1}\boldsymbol{H}_t+\mathbf{R}_t\triangleq\boldsymbol{D}_t$，因此通过$\boldsymbol{D}_t^{-1/2}\boldsymbol{I}_t$可产生服从$\mathcal{N}(\boldsymbol{0},\boldsymbol{I})$独立同分布的观测序列，即$\{\boldsymbol{I}_t\}_{t=0}^{\infty}$。其中$\boldsymbol{D}^{-1/2}$表示3.2节讨论的矩阵$\boldsymbol{D}$的平方根。增益$\boldsymbol{K}_t$可以写为$\boldsymbol{\Sigma}_{t/t-1}\boldsymbol{H}_t\boldsymbol{D}_t^{-1/2}\boldsymbol{D}_t^{-1/2}$，因此 Kalman-Bucy 滤波器实际上给出了等价于观测序列的白色序列。

例 5.2.2 **时不变单变量情况**

说明卡尔曼滤波器的最简单模型是所有参数都独立于时间的一维($m=k=1$)模型。特别地，考虑如下模型

$$
\begin{aligned}
X_{n+1} &= fX_n + U_n, \quad n = 0,1,\cdots \\
Y_n &= hX_n + V_n, \quad n = 0,1,\cdots
\end{aligned}
\tag{5.2.27}
$$

其中，$\{U_n\}_{n=0}^{\infty}$和$\{V_n\}_{n=0}^{\infty}$是独立同分布随机变量序列，分别服从$\mathcal{N}(0, q)$和$\mathcal{N}(0, r)$，且$X_0 \sim \mathcal{N}(m_0, \boldsymbol{\Sigma}_0)$，并且 f、h、q、r 和 $\boldsymbol{\Sigma}_0$ 都是标量。

这种情况下的递归估计是

$$
\hat{X}_{t+1|t} = f\,\hat{X}_{t|t}, \quad t = 0,1,\cdots \tag{5.2.28a}
$$

和

$$
\hat{X}_{t|t} = \hat{X}_{t|t-1} + K_t(Y_t - h\,\hat{X}_{t|t-1}), \quad t = 0,1,\cdots \tag{5.2.28b}
$$

且 K_t 由下式给出

$$
K_t = \frac{\boldsymbol{\Sigma}_{t|t-1}h}{h^2\boldsymbol{\Sigma}_{t|t-1} + r} = \frac{1}{h}\,\frac{\boldsymbol{\Sigma}_{t|t-1}}{\boldsymbol{\Sigma}_{t|t-1} + r/h^2} \tag{5.2.29}
$$

从式(5.2.29)的表达式中可以看出卡尔曼增益在测量更新和卡尔曼滤波器中的作用。特别地，$\Sigma_{t|t-1}$是由$\boldsymbol{Y}_0^{t-1}$ 估计 $\boldsymbol{X}_t$ 产生的最小均方误差(MSE)，并且比值 r/h^2 是观测值中噪声功率的度量。这是因为，在 $h \neq 0$ 时，$Y_t/h = X_t + V_t/h$ 等价于对 Y_t 的测量结果，且 V_t/h 的方差是 r/h^2。从这些观测结果可以发现，如果 X_t 的预测值比 X_t 的当前观测值更准确(即 $\boldsymbol{\Sigma}_{t|t-1} \ll r/h^2$)，那么增益 $K_t \approx 0$ 并且$\hat{X}_{t|t} \approx \hat{X}_{t|t-1}$。换言之，与 X_t 的当前观测值相比，X_t 的预测值更可信，因此保留其预测值。相反，如果当前观测值比预测值更准确(即$\boldsymbol{\Sigma}_{t|t-1} \gg r/h^2$)，那么卡尔曼增益 $K_t \approx 1/h$ 并且$\hat{X}_{t|t} \approx Y_t/h$，此时，预测值被忽略，而保留测量值。当然，在这两种极端情况之间，新息的结果由两者的综合平衡来决定。虽然向量不能像标量那样进行参数化，但其更新方式类似。

为比较测量更新与例 4.2.2 中讨论的贝叶斯估计的异同，现将测量更新方程改写为

$$
\hat{X}_{t|t} = \frac{v^2d^2\,\hat{\theta}_1 + \mu}{v^2d^2 + 1} \tag{5.2.30}
$$

其中，定义$\hat{\theta}_1 = Y_t/h$，$\mu = \hat{X}_{t|t-1}$，$v^2 = \boldsymbol{\Sigma}_{t|t-1}$和 $d^2 = h^2/r$。对比式(5.2.30)和式(4.2.34)可知：X_t 依赖于 Y_0^{t-1} 的条件分布可视为 X_t 的先验分布[即$\mathcal{N}(\hat{X}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1})$]，更新过程根据 v^2d^2 的值来平衡观测值$\boldsymbol{Y}_t$ 和先验数据。(这也正是命题 5.2.1 中测量更新过程的本质。)

对于这种时不变的标量模型，估计方差的时间更新和测量更新变为

$$
\boldsymbol{\Sigma}_{t+1|t} = f^2\boldsymbol{\Sigma}_{t|t} + q \tag{5.2.31a}
$$

和

$$
\boldsymbol{\Sigma}_{t|t} = \frac{\boldsymbol{\Sigma}_{t|t-1}}{\frac{h^2}{r}\boldsymbol{\Sigma}_{t|t-1} + 1} \tag{5.2.31b}
$$

注意，可以通过消除这些方程之间的耦合来得到每个变量单独的递归方程。例如，将式(5.2.31b)代入式(5.2.31a)可得递归方程(初始值是 $\boldsymbol{\Sigma}_{0|-1} = \boldsymbol{\Sigma}_0$)

$$\boldsymbol{\Sigma}_{t+1|t}=\frac{f^2\boldsymbol{\Sigma}_{t|t-1}}{h^2\boldsymbol{\Sigma}_{t|t-1}/r+1}+q,\quad t=0,1,\cdots \tag{5.2.32}$$

式(5.2.32)中存在一个问题：随着时间的增加，通过递归产生的序列是否会逼近于常量？若是，那么卡尔曼增益也会趋近于常量，并且 Kalman-Bucy 滤波器将会收敛到时不变系统。若 $\boldsymbol{\Sigma}_{t+1|t}$逼近常数$\boldsymbol{\Sigma}_{\infty}$，那么$\boldsymbol{\Sigma}_{\infty}$必须满足

$$\boldsymbol{\Sigma}_{\infty}=\frac{f^2\boldsymbol{\Sigma}_{\infty}}{h^2\boldsymbol{\Sigma}_{\infty}/r+1}+q \tag{5.2.33}$$

这是因为$\boldsymbol{\Sigma}_{t+1|t}$和$\boldsymbol{\Sigma}_{t|t-1}$都趋近于$\boldsymbol{\Sigma}_{\infty}$。显然，式(5.2.33)是二次方程，存在唯一正解

$$\boldsymbol{\Sigma}_{\infty}=\frac{1}{2}\left\{\left[\frac{r}{h^2}(1-f^2)-q\right]^2+\frac{4rq}{h^2}\right\}^{1/2}-\frac{r}{2h^2}(1-f^2)+q \tag{5.2.34}$$

结合式(5.2.32)和式(5.2.33)，得[⊖]

$$|\boldsymbol{\Sigma}_{t+1|t}-\boldsymbol{\Sigma}_{\infty}|=f^2\left|\frac{\boldsymbol{\Sigma}_{t|t-1}}{h^2\boldsymbol{\Sigma}_{t|t-1}/r+1}-\frac{\boldsymbol{\Sigma}_{\infty}}{h^2\boldsymbol{\Sigma}_{\infty}/r+1}\right|\leqslant f^2|\boldsymbol{\Sigma}_{t|t-1}-\boldsymbol{\Sigma}_{\infty}|,\quad t=0,1,\cdots \tag{5.2.35}$$

进而有下面的表达式成立

$$|\boldsymbol{\Sigma}_{t+1|t}-\boldsymbol{\Sigma}_{\infty}|\leqslant f^{2(t+1)}|\boldsymbol{\Sigma}_0-\boldsymbol{\Sigma}_{\infty}| \tag{5.2.36}$$

若$|f|<1$，那么式(5.2.36)意味着当$t\to\infty$时，$\boldsymbol{\Sigma}_{t+1|t}\to\boldsymbol{\Sigma}_{\infty}$。因此对 Kalman-Bucy 滤波器来说，条件$|f|<1$是充分的，且能够使滤波器性能达到稳态。[注意，条件$|f|<1$也是式(5.2.27)所描述的模型满足渐近稳定的条件。] ◀

例 5.2.3 跟踪扫描雷达

雷达的常用功能是定期对一些区域进行扫描(这些区域称为空域)，并在每个扫描角度进行位置测量以对不同目标进行轨迹跟踪。雷达也会对目标下一次扫描时的位置进行预测。由于目标的移动轨迹对雷达是未知的，从设计最优跟踪算法的角度出发，通常假设运动目标具有随机加速度。这类运动的一个简单模型是考虑加速度在各次扫描间是独立同分布的高斯过程。目标运动通常是三维的，简单起见，本书仅在一维空间中进行讨论。特别地，假设状态方程和量测方程如下

$$\begin{aligned}\begin{pmatrix}P_{n+1}\\V_{n+1}\end{pmatrix}&=\begin{pmatrix}1&T_s\\0&1\end{pmatrix}\begin{pmatrix}P_n\\V_n\end{pmatrix}+\begin{pmatrix}0\\T_s\end{pmatrix}A_n\\Y_n&=(1\quad 0)\begin{pmatrix}P_n\\V_n\end{pmatrix}+\varepsilon_n\end{aligned} \tag{5.2.37}$$

其中，P_n 和 V_n 分别表示第 n 次扫描时目标的位置和径向速度，T_s 是雷达完成每次扫描所需的时间，A_n 是第 n 次扫描时目标的加速度，Y_n 是第 n 次观测时的目标位置，ε_n 是测量误差。(对于三维空间的跟踪问题，需要采用六状态、三测量模型。不过，如果三维运动的加速度和测量噪声相互独立，则可以对三个维度分别进行跟踪。)

⊖ 要得到式(5.2.35)，可定义 $g(x)=x/(ax+1)$，其中 $a=h^2/r$。根据泰勒定理，对于任意实数 x 和 y，当 $x\leqslant\xi\leqslant y$ 时，有

$$|g(x)-g(y)|=|x-y||g'(\xi)|$$

当 $\xi\geqslant 0$ 时，$g'(\xi)=1/(a\xi+1)^2$ 满足 $|g'(\xi)|\leqslant 1$。因为 $\boldsymbol{\Sigma}_{t|t-1}>0$ 且 $\boldsymbol{\Sigma}_{\infty}>0$，所以

$$|g(\boldsymbol{\Sigma}_{t|t-1})-g(\boldsymbol{\Sigma}_{\infty})|\leqslant|\boldsymbol{\Sigma}_{t|t-1}-\boldsymbol{\Sigma}_{\infty}|$$

假设所有数据都服从具有时不变特性的高斯分布，则最优跟踪/预测方程为

$$\begin{bmatrix} \hat{P}_{t+1|t} \\ \hat{V}_{t+1|t} \end{bmatrix} = \begin{bmatrix} \hat{P}_{t|t} + T_s \hat{V}_{t|t} \\ \hat{V}_{t|t} \end{bmatrix} \tag{5.2.38}$$

和

$$\begin{bmatrix} \hat{P}_{t|t} \\ \hat{V}_{t|t} \end{bmatrix} = \begin{bmatrix} \hat{P}_{t|t-1} \\ \hat{V}_{t|t-1} \end{bmatrix} + \begin{bmatrix} K_{t,1} \\ K_{t,2} \end{bmatrix} (Y_t - \hat{P}_{t|t-1}) \tag{5.2.39}$$

增益矩阵 $\boldsymbol{K}_t$ 是一个 2×1 的向量。它由下式给出

$$\begin{bmatrix} K_{t,1} \\ K_{t,2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{t|t-1}(1,1)/(\boldsymbol{\Sigma}_{t|t-1}(1,1)+r) \\ \boldsymbol{\Sigma}_{t|t-1}(2,1)/(\boldsymbol{\Sigma}_{t|t-1}(1,1)+r) \end{bmatrix} \tag{5.2.40}$$

其中，$\boldsymbol{\Sigma}_{t|t-1}(k, l)$是矩阵 $\boldsymbol{\Sigma}_{t|t-1}$的第$(k-l)$个元素，$r$ 是测量噪声的方差。当然，增益矩阵可根据定理 5.2.1 通过递归计算得到。

为降低跟踪方程的计算复杂度，可用时不变滤波器代替式(5.2.39)中的时变滤波器，即

$$\begin{bmatrix} \hat{P}_{t|t} \\ \hat{V}_{t|t} \end{bmatrix} = \begin{bmatrix} \hat{P}_{t|t-1} \\ \hat{V}_{t|t-1} \end{bmatrix} + \begin{bmatrix} \alpha \\ \beta/T_s \end{bmatrix} (Y_t - \hat{P}_{t|t-1}) \tag{5.2.41}$$

其中，α 和 β 都是常数。该跟踪器也称为 α-β 跟踪器，参数 α 和 β 需要根据响应速度、跟踪精度等系统性能进行折衷选择。

“边跟踪边扫描”雷达问题将在后文中进行讨论。◀

就命题 5.2.1 中描述的一般 Kalman-Bucy 滤波器而言，在式(5.2.14)和式(5.2.16)中可以分离耦合递归关系，即可得到仅有预测$\hat{\boldsymbol{X}}_{t+1|t}$和$\boldsymbol{\Sigma}_{t+1|t}$而不包含滤波$\hat{\boldsymbol{X}}_{t|t}$和$\boldsymbol{\Sigma}_{t|t}$的递归方程，反之亦然(如例 5.2.2)。例如，将测量更新代入时间更新，可得

$$\hat{\boldsymbol{X}}_{t+1|t} = \boldsymbol{F}_t \hat{\boldsymbol{X}}_{t|t-1} + \boldsymbol{F}_t\boldsymbol{K}_t \boldsymbol{I}_t, \quad t = 0,1,\cdots \tag{5.2.42a}$$

和

$$\boldsymbol{\Sigma}_{t+1|t} = \boldsymbol{F}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{F}_t^T - \boldsymbol{F}_t\boldsymbol{K}_t\boldsymbol{H}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{F}_t^T + \boldsymbol{G}_t\boldsymbol{Q}_t\boldsymbol{G}_t, \quad t = 0,1,\cdots \tag{5.2.42b}$$

预测滤波器式(5.2.42a)是一个由新息序列驱动的线性随机系统。因此，这个系统与原跟踪系统具有相同的动态特性($\boldsymbol{F}_t$)。就跟踪状态向量$\boldsymbol{X}_t$ 而言，重建了一个与$\boldsymbol{X}_t$ 动态特性相同的系统，其中新息通过矩阵序列 $\boldsymbol{F}_t\boldsymbol{K}_t$ 影响状态向量。

方差更新方程(5.2.42b)是由状态矩阵确定的动态系统，具有非线性特性，因为右边第二部分中包含 $\boldsymbol{K}_t$ 项的部分依赖于 $\boldsymbol{\Sigma}_{t|t-1}$。该方程就是著名的(离散时间)Riccati 微分方程。正如例 5.2.2 中的标量情况，这个方程的时不变形式(即 $\boldsymbol{F}_t$、$\boldsymbol{G}_t$、$\boldsymbol{H}_t$、$\boldsymbol{Q}_t$ 和$\boldsymbol{R}_t$ 都独立于 t)会收敛到稳态。对于 $\boldsymbol{\Sigma}_{t+1|t}$而言，它收敛于稳定状态的充分条件是 $\boldsymbol{F}$ 的所有特征值都小于 1(该条件是原始系统渐近稳定的充要条件)。另外一个关键问题是，求解式(5.2.42b)中的 $\boldsymbol{K}_t$ 涉及计算逆矩阵$(\boldsymbol{H}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}_t^T+\boldsymbol{R}_t)^{-1}$，它在数值上可能不稳定。为了将 $\boldsymbol{\Sigma}_{t+1|t}$的根变得更加分散从而避免产生数值问题，有必要寻求式(5.2.42)的等效方程。Anderson 和 Moore 在 1979 年讨论了很多这方面的相关问题。

在推导 Kalman-Bucy 滤波器的过程中，利用了上述有关状态和测量模型的假设。虽然

所有这些假设都是必要的，但正如前文所述，通过重新建立适当的模型或者对一些性能指标进行重新选择后，可以避开其中的一些假设。例如，如果将噪声序列建模成线性随机系统的输出，就可以放宽输入序列$\{\boldsymbol{U}_k\}_{k=0}^{\infty}$和噪声序列$\{\boldsymbol{V}_k\}_{k=0}^{\infty}$的独立性假设。此时，综合原系统的状态向量和附加系统的状态向量，会得到一个驱动序列和测量序列独立的高维模型，进而可用标准卡尔曼滤波器求解。不过，这种系统的弊端在于它需要求解一个更高维的滤波器来实现对噪声和输入信号的跟踪。

下面结合例 5.2.4 来进一步说明扩维后卡尔曼滤波器的具体求解方法。

例 5.2.4 具有相关加速度序列的跟踪扫描雷达

重新考虑例 5.2.3 中的"边跟踪边扫描"雷达。对典型的雷达扫描速度和感兴趣的目标速度区间而言，假设"目标加速度在每次扫描之间相互独立"不太符合实际情况，例如目标惯性将使得相邻扫描的目标加速度存在相关性。描述扫描间目标加速度具有相关性的简单模型是加速度序列$\{A_n\}_{n=0}^{\infty}$按照随机序列产生，即

$$A_{n+1}=\rho A_n+W_n,\quad n=0,1,\cdots \tag{5.2.43}$$

其中，初始条件 A_0 服从高斯分布，$\{W_n\}_{n=0}^{\infty}$是独立同分布的高斯序列，ρ 是满足 $0\leqslant\rho<1$ 的参数。$\rho=0$ 意味着加速度序列之间没有相关性，ρ 越大说明加速度序列的相关性越强。

在加速度满足式(5.2.43)的条件下，式(5.2.37)所描述的模型不再满足 Kalman-Bucy 滤波器所要求的假设。然而，如果将加速度看成是状态(而不是一个输入)变量，可得到包含加速度动态特性的扩展状态模型，即

$$\begin{pmatrix}P_{n+1}\\V_{n+1}\\A_{n+1}\end{pmatrix}=\begin{pmatrix}1&T_s&0\\0&1&T_s\\0&0&\rho\end{pmatrix}\begin{pmatrix}P_n\\V_n\\A_n\end{pmatrix}+\begin{pmatrix}0\\0\\1\end{pmatrix}W_n,\quad n=0,1,\cdots \tag{5.2.44a}$$

$$Y_n=(1\quad 0\quad 0)\begin{pmatrix}P_n\\V_n\\A_n\end{pmatrix}+\varepsilon_n,\quad n=0,1,\cdots \tag{5.2.44b}$$

进而可得到状态向量的递归估计为

$$\begin{pmatrix}\hat{P}_{t+1|t}\\\hat{V}_{t+1|t}\\\hat{A}_{t+1|t}\end{pmatrix}=\begin{pmatrix}\hat{P}_{t|t}+T_s\hat{V}_{t|t}\\\hat{V}_{t|t}+T_s\hat{A}_{t|t}\\\rho\hat{A}_{t|t}\end{pmatrix} \tag{5.2.45a}$$

和

$$\begin{pmatrix}\hat{P}_{t|t}\\\hat{V}_{t|t}\\\hat{A}_{t|t}\end{pmatrix}=\begin{pmatrix}\hat{P}_{t|t-1}\\\hat{V}_{t|t-1}\\\hat{A}_{t|t-1}\end{pmatrix}+\begin{pmatrix}K_{t,1}\\K_{t,2}\\K_{t,3}\end{pmatrix}(Y_t-\hat{P}_{t|t-1}) \tag{5.2.45b}$$

增益为

$$\begin{pmatrix}K_{t,1}\\K_{t,2}\\K_{t,3}\end{pmatrix}=\begin{pmatrix}\boldsymbol{\Sigma}_{t|t-1}(1,1)/(\boldsymbol{\Sigma}_{t|t-1}(1,1)+r)\\\boldsymbol{\Sigma}_{t|t-1}(2,1)/(\boldsymbol{\Sigma}_{t|t-1}(1,1)+r)\\\boldsymbol{\Sigma}_{t|t-1}(3,1)/(\boldsymbol{\Sigma}_{t|t-1}(1,1)+r)\end{pmatrix} \tag{5.2.46}$$

现在要跟踪的不仅仅是目标的位置和径向速度，还需要跟踪目标的加速度。与例5.2.3中的低阶模型类似，式(5.2.45b)中的增益向量在实际中也可用常向量替换，即

$$\begin{pmatrix} \alpha \\ \beta/T_s \\ \gamma/T_s^2 \end{pmatrix}$$

这就是著名的α-β-γ跟踪器，可通过选择参数α、β和γ的值来达到预期的性能。 ◀

上述例子说明了在卡尔曼滤波器中如何处理输入序列的独立性问题。关于独立性问题更详细的阐述，可以参考Anderson和Moore(1979)的著作。卡尔曼滤波器的另一关键假设是状态和测量方程具有线性和高斯特性。当然，如果仅在线性滤波器中寻求最优性能，那么可忽略高斯假设。Kalman-Bucy滤波器由二阶统计量(均值和方差)确定，因此对于任意初始条件、输入和噪声序列而言，无论是否服从高斯分布，只要存在二阶统计量，Kalman-Bucy滤波器在所有线性滤波器中是最优的，相关问题将在5.3节中进行讨论。相比高斯分布假设条件，状态和观测方程的线性假设很难放宽。事实上，如果没有线性假设，在最小均方误差(MMSE)意义上求解状态估计值将非常困难。非线性系统的一些状态估计方法将在7.3节关于连续信号估计的内容中介绍。

5.3 线性估计

在5.2节中讨论了求解具有高斯分布特性的线性随机系统的最优估计器问题。卡尔曼滤波器不仅对高斯线性模型是最优的，而且对于具有二阶统计量(均值和方差)的非高斯线性模型也是最优的。事实上，非高斯模型线性估计器是最优线性估计器的特例，它只需要二阶统计量即可确定估计器的具体形式。本节将继续探讨该估计器，并将运用相关理论解决Wiener-Kolmogorov滤波问题。

假设有两个随机序列$\{Y_n\}_{n=-\infty}^{\infty}$和$\{X_n\}_{n=-\infty}^{\infty}$，在$a \leqslant n \leqslant b$期间对$Y_n$进行观测，并利用这些观测来估计$X_t$。当然，条件均值$\hat{X}_t = E\{X_t \mid Y_a^b\}$是$X_t$(在MMSE意义上)的最优估计。不过，如果观测的数量$b-a+1$很大，除非这个问题的结构很特殊(比如Kalman-Bucy模型)，否则计算条件均值会相当复杂。此外，确定条件均值需要知道变量X_t，Y_a，…，Y_b的联合分布，而这在实际中通常难以准确获取。

对估计器施加线性约束并求解该约束下的最小均方误差估计(MSE)是降低运算复杂度的有效方式，例如考虑形式如下的估计值$\hat{X}_t$

$$\hat{X}_t = \sum_{n=a}^{b} h_{t,n} Y_n + c_t \tag{5.3.1}$$

其中，$h_{t,a}$，…，$h_{t,b}$和c_t均为标量[⊖]。此外，在线性约束下求解最优估计器仅需要已知二阶统计特性，避免了求解随机变量的联合分布。在分析最优估计器之前，首先讨论式(5.3.1)求和的一些性质。

⊖ 由于存在常数项c_t，式(5.3.1)的形式并非线性变换，更确切地说，它应该是仿射变换。为标准起见，仍采用“线性变换”。值得注意的是，如果X_t，Y_a，…，Y_a是联合高斯分布随机变量，那么$E\{X_t \mid Y_a^b\}$满足式(5.3.1)，此时将得到全局最优估计器。

虽然对于有限的 a 和 b，式(5.3.1)求和的意义非常明确，但是需要考虑 $a=-\infty$、$b=+\infty$或二者同时趋向无穷的情况。换言之，从实践的观点来说，式(5.3.1)的意义是很明确的，但是从理论分析的角度，必须明确界定随机变量求和的意义。本书采用均方和定义，即对于 $a=-\infty$、b 有限的情况，式(5.3.1)意味着

$$\lim_{m\to-\infty} E\left\{\left(\sum_{n=m}^{b} h_{t,n}Y_n + c_t - \hat{X}_t\right)^2\right\} = 0 \tag{5.3.2}$$

式(5.3.1)求和对于 $b=+\infty$、a 有限和 $a=-\infty$、$b=+\infty$的情况同样适用。根据式(5.3.2)的极限定义，当 $a=-\infty$、b 有限时，观测序列应理解为 $a<t\leqslant b$ 而不是 $a\leqslant t\leqslant b$；当 $b=+\infty$时，对观测序列的理解与此类似。

为进一步讨论线性估计问题，本节假设$\{X_n\}_{n=-\infty}^{\infty}$和$\{Y_n\}_{n=-\infty}^{\infty}$都是二阶序列，即有 $E\{X_n^2\}<\infty$和 $E\{Y_n^2\}<\infty$成立。同时，记 H_a^b 表示在 Y_a^b 基础上满足式(5.3.1)的所有估计值的集合，下面首先给出与 H_a^b 有关的结论，这在本书后续讨论中会频繁使用。

命题 5.3.1 假设$\hat{X}_t\in H_a^b$，那么：

1) $E\{(\hat{X}_t)^2\}<\infty$；

2) 如果 Z 是满足 $E\{Z^2\}<\infty$的随机变量，则

$$E\{Z\hat{X}_t\} = \sum_{n=a}^{b} h_{t,n}E\{ZY_n\} + c_tE\{Z\}$$

证明 如果 a、b 都是有限值，那么这两个性质显然成立。此时，证明性质 1)需要利用不等式$(x+y)^2\leqslant 4(x^2+y^2)$，性质 2)由期望的线性性质直接得到。为了证明 a、b 有一个或者两个趋于无穷时也成立，考察 $a=-\infty$、b 有限的情况(其他两种情形的证明过程类似)。

为了证明性质 1)，考虑 $m<b$，有

$$\hat{X}_t = \sum_{n=m}^{b} h_{t,n}Y_n + c_t + \left(\hat{X}_t - \sum_{n=m}^{b} h_{t,n}Y_n - c_t\right) \tag{5.3.3}$$

使用不等式$(x+y)^2\leqslant 4(x^2+y^2)$并求期望，得

$$E\{(\hat{X}_t)^2\} \leqslant 4E\left\{\left(\sum_{n=m}^{b} h_{t,n}Y_n + c_t\right)^2\right\} + 4E\left\{\left(\hat{X}_t - \sum_{n=m}^{b} h_{t,n}Y_n - c_t\right)^2\right\} \tag{5.3.4}$$

对于有限的 a 和 b，式(5.3.4)右边第一项的值也是有限的，右边第二项在 m 趋于无穷时收敛到 0。因此，一定存在 m，使得第二项为有限值，即 $E\{(\hat{X}_t)^2\}<\infty$成立。

为了证明性质 2)，在 $m<b$ 时有

$$E\{Z\hat{X}_t\} - \sum_{n=m}^{b} h_{t,n}E\{ZY_n\} - c_tE\{Z\} = E\left\{Z\left(\hat{X}_t - \sum_{n=m}^{b} h_{t,n}Y_n - c_t\right)\right\} \tag{5.3.5}$$

利用施瓦兹不等式，得

$$\left|E\left\{Z\left(\hat{X}_t - \sum_{n=m}^{b} h_{t,n}Y_n - c_t\right)\right\}\right|^2 \leqslant E\{Z^2\}E\left\{\left(\hat{X}_t - \sum_{n=m}^{b} h_{t,n}Y_n - c_t\right)^2\right\} \tag{5.3.6}$$

结合假设 $E\{Z^2\}<\infty$并根据定义有 $m\to-\infty$时 $E\left\{\left(\hat{X}_t - \sum_{n=m}^{b} h_{t,n}Y_n - c_t\right)^2\right\}\to 0$，综合式(5.3.5)和式(5.3.6)后可证明性质 2)成立。至此完成命题证明。 □

一旦将估计器约束成为式(5.3.1)的形式，则 MMSE 意义下的最优估计器可表示为

$$\min_{\hat{X}_t \in H_a^b} E\{(\hat{X}_t - X_t)^2\} \tag{5.3.7}$$

在下面的命题中给出了上述最优估计器的解。

命题 5.3.2　正交原理

$\hat{X}_t \in H_a^b$ 是(5.3.7)的解，当且仅当

$$E\{(\hat{X}_t - X_t)Z\} = 0, \quad Z \in H_a^b \tag{5.3.8}$$

证明　假设$\hat{X}_t$ 满足式(5.3.8)，且$\widetilde{X}_t$ 为 H_a^b 集合内的任意不同于$\hat{X}_t$ 的估计，则$\widetilde{X}_t$ 的 MSE 为

$$\begin{aligned} E\{(X_t - \widetilde{X}_t)^2\} &= E\{(X_t - \hat{X}_t + \hat{X}_t - \widetilde{X}_t)^2\} \\ &= E\{(X_t - \hat{X}_t)^2\} + 2E\{(X_t - \hat{X}_t)(\hat{X}_t - \widetilde{X}_t)\} + E\{(\hat{X}_t - \widetilde{X}_t)^2\} \end{aligned} \tag{5.3.9}$$

易知$\hat{X}_t \in H_a^b$ 和$\widetilde{X}_t \in H_a^b$，这意味着$(\hat{X}_t - \widetilde{X}_t) \in H_a^b$。假设式(5.3.8)成立，则式(5.3.9)右边第二项为 0，进而有

$$E\{(X_t - \widetilde{X}_t)^2\} = E\{(X_t - \hat{X}_t)^2\} + E\{(\hat{X}_t - \widetilde{X}_t)^2\} \geqslant E\{(X_t - \hat{X}_t)^2\} \tag{5.3.10}$$

既然式(5.3.10)中的$\widetilde{X}_t$ 是任意的，因此式(5.3.8)是式(5.3.7)的充分条件。

为证明式(5.3.8)的必要性，假设$\widetilde{X}_t \in H_a^b$，同时存在一个 $Z \in H_a^b$ 满足 $E\{(X_t - \widetilde{X}_t)Z\} \neq 0$。定义估计值$\hat{X}_t$ 为

$$\hat{X}_t = \widetilde{X}_t + \frac{E\{(X_t - \widetilde{X}_t)Z\}}{E\{Z^2\}} Z \tag{5.3.11}$$

(条件 $E\{(X_t - \widetilde{X}_t)Z\} \neq 0$ 意味着 $E\{Z^2\} > 0$)经过计算后可得

$$E\{(X_t - \hat{X}_t)^2\} = E\{(X_t - \widetilde{X}_t)^2\} - \frac{|E\{(X_t - \widetilde{X}_t)Z\}|^2}{EZ^2} < E\{(X_t - \widetilde{X}_t)^2\} \tag{5.3.12}$$

因此，$\hat{X}_t$ 是比$\widetilde{X}_t$ 更好的估计，即$\widetilde{X}_t$ 不满足式(5.3.7)。换言之，式(5.3.8)是最优估计器的必要条件。至此，完成了命题的证明过程。　□

命题 5.3.2 表明：在给定 Y_a^b 的条件下，当且仅当估计误差 $X_t - \hat{X}_t$ 与观测 Y_a^b 的任意线性变换均正交时，$\hat{X}_t$ 是 X_t 的线性 MMSE 估计。这就是著名的正交原理[⊖]，它是投影定理的特例。在有限维向量空间中，投影定理具有更一般的形式。

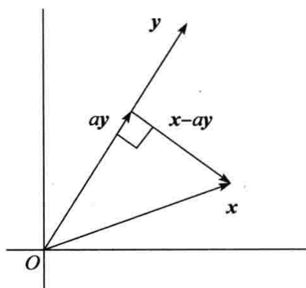


图 5.3.1　正交原理说明

假设 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 维数相同，并用常量 α 乘以 $\boldsymbol{y}$ 来近似 $\boldsymbol{x}$，以使得误差向量 $\boldsymbol{x} - \alpha\boldsymbol{y}$ 的长度尽可能小。显然，当且仅当误差向量垂直于(即正交于)沿 $\boldsymbol{y}$ 的直线(参见图 5.3.1)时，α 可最小化误差向量。此时，$\boldsymbol{x}$ 的近似估计也就是 $\boldsymbol{x}$ 在 $\boldsymbol{y}$ 方向上的投影。

⊖ 有意思的是，条件均值估计器 $\hat{X}_t = E\{X_t \mid Y_a^b\}$ 恰恰满足

$$E\{(X_t - \hat{X}_t)Z\} = 0, \quad Z \in G_a^b$$

其中 G_a^b 表示形式为 $g(Y_a^b)$ 并满足 $E\{g^2(Y_a^b)\} < \infty$ 的随机变量的集合。

类比投影问题与线性 MMSE 估计问题是非常直观的：向量 $\boldsymbol{y}$ 类似于观测 Y_a^b，沿 $\boldsymbol{y}$ 方向的向量类似于所有线性估计器的集合 H_a^b，向量 $\boldsymbol{x}$ 对应于待估计量 X_t，误差向量的长度 $\|\boldsymbol{x}-\alpha\boldsymbol{y}\|^2$ 类似于 MSE，即 $E\{(X_t-\hat{X}_t)^2\}$。综上分析，可以认为线性 MMSE 就是 X_t 在观测 Y_a^b 上的投影。

命题 5.3.2 描述了满足式(5.3.7)的解应具有的性质。下面给出更利于求解的形式。

命题 5.3.3　等价正交条件

$\hat{X}_t$ 满足式(5.3.7)当且仅当

$$E\{\hat{X}_t\}=E\{X_t\} \tag{5.3.13}$$

和

$$E\{(X_t-\hat{X}_t)Y_l\}=0,\quad a\leqslant l\leqslant b \tag{5.3.14}$$

证明　根据正交原理，分别令 $Z=1$ 和 $Z=Y_l$，即可完成式(5.3.13)和式(5.3.14)的必要性证明。证明充分性需要利用命题 5.3.1 的性质 2)。特别地，假设 $\hat{X}_t$ 满足式(5.3.13)和式(5.3.14)，并令 $Z=\sum_{n=a}^{b}h_{t,n}Y_n+c_t$，有

$$E\{(X_t-\hat{X}_t)Z\}=\sum_{n=a}^{b}h_{t,n}E\{(X_t-\hat{X}_t)Y_n\}+c_tE\{X_t-\hat{X}_t\}=0 \tag{5.3.15}$$

即式(5.3.8)成立。至此完成证明过程。　□

利用命题 5.3.2，可得到确定式(5.3.1)中最优滤波器系数的方程。将式(5.3.1)代入式(5.3.13)，可得

$$E\left\{\sum_{n=a}^{b}h_{t,n}Y_n+c_t\right\}=E\{X_t\}$$

应用命题 5.3.1 的性质 2)，取 $Z=1$ 后，有

$$c_t=E\{X_t\}-\sum_{n=a}^{b}h_{t,n}E\{Y_n\} \tag{5.3.16}$$

结合式(5.3.1)和式(5.3.14)，可得如下关系

$$E\left\{\left(X_t-\sum_{n=a}^{b}h_{t,n}Y_n-c_t\right)Y_l\right\}=0,\quad a\leqslant l\leqslant b \tag{5.3.17}$$

将式(5.3.16)代入式(5.3.17)，可得

$$E\{[(X_t-E\{X_t\})-\sum_{n=a}^{b}h_{t,n}(Y_n-E\{Y_n\})]Y_l\}=0,\quad a\leqslant l\leqslant b$$

$$E\{(X_t-E\{X_t\})Y_l\}=\sum_{n=a}^{b}h_{t,n}E\{(Y_n-E\{Y_n\})Y_l\},\quad a\leqslant l\leqslant b$$

$$\operatorname{cov}(X_t,Y_l)=\sum_{n=a}^{b}h_{t,n}\operatorname{cov}(Y_n,Y_l),\quad a\leqslant l\leqslant b$$

最终得到

$$C_{XY}(t,l)=\sum_{n=a}^{b}h_{t,n}C_Y(n,l),\quad a\leqslant l\leqslant b \tag{5.3.18}$$

其中 $C_{XY}(t,\ l)\triangleq\operatorname{cov}(X_t,\ Y_l)$ 是序列 $\{X_n\}_{n=-\infty}^{\infty}$ 与 $\{Y_n\}_{n=-\infty}^{\infty}$ 的互相关函数，而 $C_Y(n,\ l)\triangleq\operatorname{cov}(Y_n,\ Y_l)$ 是序列 $\{Y_n\}_{n=-\infty}^{\infty}$ 的自相关函数。

式(5.3.16)和式(5.3.18)给出了确定系数$\{h_{t,n}\}_{n=a}^{b}$和c_t的充要方程，由此可得由Y_a^b估计X_t的最优线性估计器。注意，这些方程仅涉及Y_a^b和X_t的均值、自相关和互相关(二阶统计量)。与需要已知Y_a^b和X_t的联合分布的条件均值估计器$E\{X_t|Y_a^b\}$相比，线性MMSE估计器具有明显优势，这是因为二阶统计量不但比多元分布函数更容易分析建模，而且更容易从观测序列中准确估计。如果$\{h_{t,n}\}_{n=a}^{b}$满足式(5.3.18)，那么最优的c_t可由式(5.3.16)得到。事实上，c_t的作用是将$\hat{X}_t$的均值调整到与X_t相同，即最优线性估计器通常具有如下形式

$$\hat{X}_t = E\{X_t\} + \sum_{n=a}^{b} h_{t,n}(Y_n - E\{Y_n\}) \tag{5.3.19}$$

因此，为了不失一般性，可假设$\{X_n\}_{n=-\infty}^{\infty}$和$\{Y_n\}_{n=-\infty}^{\infty}$的均值为0。在此假设下有$c_t=0$，即不需要求解式(5.3.16)。

事实上，式(5.3.18)就是著名的Wiener-Hopf方程。对于有限的a、b，在理论上很容易求解该方程。特别地，式(5.3.18)是一系列($b-a+1$个)线性方程的集合，可以重新写成矩阵形式

$$\boldsymbol{\sigma}_{XY}(t) = \boldsymbol{\Sigma}_Y \boldsymbol{h}_t \tag{5.3.20}$$

其中，$\boldsymbol{\sigma}_{XY}(t) \triangleq [C_{XY}(t, a), \cdots, C_{XY}(t, \mathrm{b})]^{\mathrm{T}}$，$\boldsymbol{h}_t \triangleq [h_{t,a}, \cdots, h_{t,b}]^{\mathrm{T}}$，$\boldsymbol{\Sigma}_Y$是向量$(Y_a, \cdots, Y_{\mathrm{b}})^{\mathrm{T}}$的协方差矩阵。假设$\boldsymbol{\Sigma}_Y$是正定矩阵[⊖]，由式(5.3.20)可得最优估计器的系数为

$$\boldsymbol{h}_t = \boldsymbol{\Sigma}_Y^{-1} \boldsymbol{\sigma}_{XY}(t) \tag{5.3.21}$$

显然，对于有限的a和b，在理论上可求出MMSE估计。不过，在实际中，求解最优估计器的系数通常较为困难，这是因为矩阵$\boldsymbol{\Sigma}_Y$的求逆运算非常复杂。一般地，一个$k\times k$矩阵求逆需要k^3次操作，这里k等于观测的次数。在多数信号估计问题中，k随时间线性增长，难以实时计算式(5.3.21)的最优估计器系数。因此，对线性信号估计器的研究主要集中在探索能够有效求解最优估计器系数的特殊模型上。5.2节中Kalman-Bucy滤波器就是降低运算量的特殊模型。此外，Levinson模型和Wiener-Kolmogorov模型也是降低运算复杂度的典型模型，其中Levinson模型具有求解最优滤波系数的高效算法，而Wiener-Kolmogorov模型通过使a趋向负无穷来降低计算复杂度。下面讨论Levinson模型，Wiener-Kolmogorov模型将在5.4节中讨论。

例 5.3.1　Levinson 滤波

Levinson滤波仅考虑在时域具有平稳特性的随机序列的一步预测问题。特别地，如果观测序列$\{Y_n\}_{n=-\infty}^{\infty}$的自相关函数满足

$$C_Y(n,l) = C_Y(n-l,0) \tag{5.3.22}$$

则称该序列是相关平稳或广义平稳(w. s. s)序列。为了方便起见，省略式(5.3.22)右边第二个参数0后，可将广义平稳序列的自相关函数写成时间差的单变量函数，即

$$C_Y(n,l) \equiv C_Y(n-l) \tag{5.3.23}$$

又因为$\mathrm{cov}(Y_n, Y_l)=\mathrm{cov}(Y_l, Y_n)$，所以函数$C_Y$具有对称性，即$C_Y(n-l)=C_Y(l-n)$。

⊖ 相关矩阵$\boldsymbol{\Sigma}_Y$至少是非负定的，并非一定要求正定，这意味着存在冗余观测(如3.2节所述)。此时式(5.3.20)的解不唯一。

在 Levinson 滤波问题中，需要根据 $0\leqslant n\leqslant t$ 时刻的观测 Y_n 来预测 Y_{t+1}。采用之前的符号，有 $a=0$、$b=t$ 和 $X_t=Y_{t+1}$。显然有 $\{X_n\}_{n=-\infty}^{\infty}\equiv\{Y_{n+1}\}_{n=-\infty}^{\infty}$，$\{Y_n\}_{n=-\infty}^{\infty}$ 的互相关函数为

$$C_{XY}(t,l)=\mathrm{cov}(X_t,Y_l)=\mathrm{cov}(Y_{t+1},Y_l)=C_Y(t+1-l) \tag{5.3.24}$$

类似地，$\boldsymbol{\Sigma}_Y$ 的第$(n-l)$个元素是 $C_Y(n-l)$。因此，式(5.3.20)的 Wiener-Hopf 方程变成

$$\begin{pmatrix} C_Y(t+1) \\ C_Y(t) \\ \vdots \\ C_Y(1) \end{pmatrix}=\begin{pmatrix} C_Y(0) & C_Y(1) & \cdots & C_Y(t) \\ C_Y(1) & C_Y(0) & \cdots & C_Y(t-1) \\ \vdots & \vdots & & \vdots \\ & & & C_Y(1) \\ C_Y(t) & \cdots & C_Y(1) & C_Y(0) \end{pmatrix}\begin{pmatrix} h_{t,0} \\ h_{t,1} \\ \vdots \\ h_{t,t} \end{pmatrix} \tag{5.3.25}$$

上述方程组也称为 Yule-Walker 方程。

既然$\{Y_n\}_{n=-\infty}^{\infty}$是广义平稳的随机过程，式(5.3.25)右边的矩阵 $\boldsymbol{\Sigma}_Y$ 是 Toeplitz 矩阵，这意味着其对角线的元素都是常数[因为 $C_Y(n,\ l)=C_Y(n-l)$]。不同于一般的协方差矩阵，$k\times k$ 阶 Toeplitz 矩阵的求逆过程只需要 k^2 次操作就可实现。(1964 年 Trench 给出了 Toeplitz 矩阵的高效求逆算法)。换言之，求解广义平稳观测序列的线性 MMSE 问题的计算复杂度降低了 k 倍。就 Levinson 问题而言，利用该模型的特殊结构还可以进一步简化求解估计器系数的计算过程。注意，式(5.3.25)中左边的向量$\boldsymbol{\sigma}_{XY}(t)$与 $\boldsymbol{\Sigma}_Y$ 的第一行延迟一个时间单位后的形式相似，即可以在时域采用递归形式计算滤波器系数。

将预测值$\hat{Y}_{t+1}=\sum_{n=0}^{t}h_{t,n}Y_n$ 写成$\hat{Y}_{t+1}=-\sum_{n=0}^{t}a_{t+1,t+1-n}Y_n$ 后，系数 $a_{t,1}$，…，$a_{t,t}$可以在时间上根据如下算法(即 Levinson 算法)进行递归更新

$$a_{t+1,k}=a_{t,k}-k_t a_{t,t+1-k},\quad k=1,\cdots,t \tag{5.3.26}$$

和

$$a_{t+1,t+1}=-k_t \tag{5.3.27}$$

其中 k_t 和 $\varepsilon_t\triangleq E\{(Y_t-\hat{Y}_t)^2\}$的递归计算式为

$$\varepsilon_{t+1}=(1-k_t^2)\varepsilon_t \tag{5.3.28}$$

和

$$k_t=\frac{C_Y(t+1)+\sum_{k=1}^{t}a_{t,k}C_Y(t+1-k)}{\varepsilon_t} \tag{5.3.29}$$

初始值为 $k_0=-C_Y(1)/C_Y(0)$和 $\varepsilon_0=C_Y(0)$。注意，该算法在完成 MMSE 最优系数计算的同时求出了 ε_t 和系数 k_1，k_2，…，其中 k_1，k_2，…称为部分相关系数或者反射系数，它们对于格型滤波器完成单步预测非常有用。1984 年，Honig 和 Messerschmitt 推导了 Levinson 算法，并给出了单步预测器的具体实现结果和典型应用实例。 ◀

线性估计问题可以直接推广到 k 维观测序列的 m 维向量估计情况。考虑如下估计器

$$\hat{\boldsymbol{X}}_t=\sum_{n=a}^{b}\boldsymbol{H}_{t,n}\boldsymbol{Y}_n+\boldsymbol{c}_t \tag{5.3.30}$$

其中，$\{\boldsymbol{H}_{t,n}\}_{n=a}^{b}$是一个 $m\times k$ 维的矩阵组成的序列，$\boldsymbol{c}_t\in\mathbf{R}^m$。式(5.3.30)中，在 a 或 b 无限时，可按照均值范数定义。如当 $a=-\infty$且 b 有限时，有

$$\lim_{j\to-\infty} E\left\{\left\|\sum_{n=j}^{b}\boldsymbol{H}_{t,n}\boldsymbol{Y}_n+\boldsymbol{c}_t-\hat{\boldsymbol{X}}_t\right\|^2\right\}=0$$

其中，$\|\boldsymbol{x}\|^2\triangleq\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$。若选择一个估计值，使得

$$\min_{\hat{\boldsymbol{X}}\in\boldsymbol{H}_a^b} E\{\|\boldsymbol{X}_t-\hat{\boldsymbol{X}}_t\|^2\}$$

如前定义，$\boldsymbol{H}_a^b$ 是具有式(5.3.30)形式的估计器解的集合。根据定理5.3.2，当且仅当下式成立时，$\hat{\boldsymbol{X}}_t$ 是最优线性估计值

$$E\{(\boldsymbol{X}_t-\hat{\boldsymbol{X}}_t)^{\mathrm{T}}\boldsymbol{Z}\}=0,\quad \boldsymbol{Z}\in\boldsymbol{H}_a^b \tag{5.3.31}$$

式(5.3.31)可以等价转换为

$$E\{\hat{\boldsymbol{X}}_t\}=E\{\boldsymbol{X}_t\} \tag{5.3.32a}$$

和

$$E\{(\boldsymbol{X}_t-\hat{\boldsymbol{X}}_t)\boldsymbol{Y}_l^T\}=\mathbf{0},\quad a\leqslant l\leqslant b \tag{5.3.32b}$$

其中 $\mathbf{0}$ 表示矩阵的所有元素都是0。上式给出了求解 $\boldsymbol{c}_t$ 的最优估计值的计算方程，对应的Wiener-Hopf方程为

$$\boldsymbol{C}_{XY}(t,l)=\sum_{n=a}^{b}\boldsymbol{H}_{t,n}\boldsymbol{C}_Y(n,l),\quad a\leqslant l\leqslant b \tag{5.3.33}$$

其中，$\boldsymbol{C}_{XY}(t,\ l)\triangleq\mathrm{cov}(\boldsymbol{X}_t,\ \boldsymbol{Y}_l)$是$\{\boldsymbol{X}_n\}_{n=-\infty}^{\infty}$和$\{\boldsymbol{Y}_n\}_{n=-\infty}^{\infty}$的协方差函数，$\boldsymbol{C}_Y(n,\ l)\triangleq\mathrm{cov}(\boldsymbol{Y}_n,\ \boldsymbol{Y}_l)$，$\boldsymbol{C}_{XY}(t,\ l)$和$\boldsymbol{C}_Y(n,\ l)$分别是 $m\times k$ 和 $k\times k$ 维的矩阵。

对于有限的 a 和 b，式(5.3.33)中的Wiener-Hopf方程给出了一系列$[(b-a+1)\times m\times k]$线性方程。考虑到$(b-a+1)k$ 维向量$(\boldsymbol{Y}_a^T,\ \boldsymbol{Y}_{a+1}^T,\ \cdots,\ \boldsymbol{Y}_b^T)^{\mathrm{T}}$ 的协方差矩阵为正定矩阵，可以利用矩阵求逆求解上述方程。事实上，最小化范数的均值误差 $E\{\|\boldsymbol{X}_t-\hat{\boldsymbol{X}}_t\|^2\}$等价于最小化 $\boldsymbol{X}_t$ 的每个元素的MSE。因此，向量Wiener-Hopf方程是一系列的标量Wiener-Hopf方程，每个标量Wiener-Hopf方程含有$(b-a+1)k$ 个观测值。需要说明的是，这种结构并没有简化求解，因为观测序列的维数大大增加了计算复杂度。在标量情况下，对这些问题的研究主要集中在计算复杂度上。Levinson问题既可以写成向量形式也可以写成标量形成，其求解方法类似于例5.3.1。由此可见，Kalman-Bucy滤波器不仅是线性高斯模型下的全局MMSE估计器，而且是少量约束条件下的线性MMSE估计器。作为总结，下面结合例子来说明Kalman-Bucy滤波器是线性MMSE估计器。

例5.3.2　Kalman-Bucy滤波器是线性MMSE估计器

考虑如下线性随机系统模型

$$\boldsymbol{X}_{n+1}=\boldsymbol{F}_n\boldsymbol{X}_n+\boldsymbol{G}_n\boldsymbol{U}_n,\quad n=0,1,\cdots \tag{5.3.34a}$$

$$\boldsymbol{Y}_n=\boldsymbol{H}_n\boldsymbol{X}_n+\boldsymbol{V}_n,\quad n=0,1,\cdots \tag{5.3.34b}$$

对于每个 $n\geqslant0$，$\boldsymbol{X}_n$、$\boldsymbol{U}_n$、$\boldsymbol{Y}_n$ 和 $\boldsymbol{V}_n$ 分别是维数为 m、s、k 和 k 的随机向量，$\boldsymbol{F}_n$、$\boldsymbol{G}_n$ 和 $\boldsymbol{H}_n$ 是相应维数的矩阵。假定$\{\boldsymbol{U}_n\}_{n=-\infty}^{\infty}$和$\{\boldsymbol{V}_n\}_{n=-\infty}^{\infty}$是由零均值不相关随机向量组成的序列[即对任意 n 和 l，有 $\mathrm{cov}(\boldsymbol{V}_n,\ \boldsymbol{U}_l)=\mathbf{0}$；对 $n\neq l$，有 $\mathrm{cov}(\boldsymbol{U}_n,\ \boldsymbol{U}_l)=\mathrm{cov}(\boldsymbol{V}_n,\ \boldsymbol{V}_l)=\mathbf{0}$，$\mathbf{0}$ 表示全零矩阵]。初始条件 $\boldsymbol{X}_0$ 与$\{\boldsymbol{U}_n\}_{n=-\infty}^{\infty}$和$\{\boldsymbol{V}_n\}_{n=-\infty}^{\infty}$均无关。同时假定 $\boldsymbol{U}_n$ 和 $\boldsymbol{V}_n$ 的协方差矩阵 $\boldsymbol{Q}_n$ 和 $\boldsymbol{R}_n$ 已知，且已知 $\boldsymbol{X}_0$ 的均值 $\boldsymbol{m}_0$ 和方差矩阵 $\boldsymbol{\Sigma}_0$。除去上述假设之外，随机变量具有任意统计特性(注意，这里并未假设服从高斯统计特性)。

遵循上述假设，定理 5.2.1 中的 Kalman-Bucy 递归滤波器是由观测 $\boldsymbol{Y}_0^t$ 中的 $\boldsymbol{X}_t$ 和 $\boldsymbol{X}_{t+1}$ 构成的线性 MMSE 估计器，在此不再详细展开[⊖]。正交性原理可用来推导时间更新估计器 $\hat{\boldsymbol{X}}_{t+1|t}=\boldsymbol{F}_t\hat{\boldsymbol{X}}_{t|t}$。特别地，在给定 $\boldsymbol{Y}_0^t$ 的情况下，假设 $\hat{\boldsymbol{X}}_{t|t}$ 是 $\boldsymbol{X}_t$ 的最优线性估计。对 $\boldsymbol{Z}\in\boldsymbol{H}_0^t$，考虑

$$E\{(\boldsymbol{X}_{t+1}-\boldsymbol{F}_t\hat{\boldsymbol{X}}_{t|t})^{\mathrm{T}}\boldsymbol{Z}\} \tag{5.3.35}$$

代入式(5.3.34a)后，式(5.3.35)可写成

$$E\{(\boldsymbol{F}_t\boldsymbol{X}_t+\boldsymbol{G}_t\boldsymbol{U}_t-\boldsymbol{F}_t\hat{\boldsymbol{X}}_{t|t})^{\mathrm{T}}\boldsymbol{Z}\}=E\{(\boldsymbol{X}_t-\hat{\boldsymbol{X}}_{t|t})^{\mathrm{T}}\boldsymbol{F}_t^T\boldsymbol{Z}\}+E\{\boldsymbol{U}^{\mathrm{T}}\boldsymbol{Z}\}\boldsymbol{G}_t^T \tag{5.3.36}$$

由于假定 $\hat{\boldsymbol{X}}_{t|t}$ 是利用 $\boldsymbol{Y}_0^t$ 估计 $\boldsymbol{X}_t$ 的最优线性估计，并且对于 $\boldsymbol{Z}\in\boldsymbol{H}_0^t$ 有 $\boldsymbol{F}_t^T\boldsymbol{Z}\in\boldsymbol{H}_0^t$。式(5.3.36)的右边第一项在 $\boldsymbol{Z}\in\boldsymbol{H}_0^t$ 时恒等于 0。观察式(5.3.36)右边第二项，$\boldsymbol{Z}$ 是 $\boldsymbol{Y}_0^t$ 的线性变换，而 $\boldsymbol{Y}_0^t$ 又是 $\boldsymbol{X}_0$、$\boldsymbol{U}_0^{t-1}$ 和 $\boldsymbol{V}_0^t$ 的线性变换，它们与 $\boldsymbol{U}_t$ 都不相关，即 $\boldsymbol{U}_t$ 和 $\boldsymbol{Z}$ 不相关且对于 $\boldsymbol{Z}\in\boldsymbol{H}_0^t$ 有 $E\{(\boldsymbol{X}_{t+1}-\boldsymbol{F}_t\hat{\boldsymbol{X}}_{t|t})^{\mathrm{T}}\boldsymbol{Z}\}=0$ 成立。显然，基于正交原理，在给定 $\boldsymbol{Y}_0^t$ 条件下，$\boldsymbol{F}_t\hat{\boldsymbol{X}}_{t|t}$ 是 $\boldsymbol{X}_{t+1}$ 的最优线性估计器，其时间更新协方差的证明过程与命题 5.2.1 中的高斯分布情形基本相同。 ◀

5.4 Wiener-Kolmogorov 滤波

在 5.3 节中，推导了 Wiener-Hopf 方程，该方程通过对一组随机变量 Y_a，…，Y_b 的观测，确定对随机变量 X_t 进行最优线性估计时的相关参数。在大部分信号估计的实际应用中，观测样本的个数 $b-a+1$ 随 t 线性增长，因此，需要给出进一步假设，以便更有效地计算估计值的相关参数。其中，在 Levinson 和 Kalman-Bucy 滤波模型中分别给出了两组假设，在 Wiener-Kolmogorov 模型中给出了另一组简化的假设，可用于求解最优线性估计问题，并在信号估计领域得到了广泛应用。本节将对 Wiener-Kolmogorov 模型进行详细介绍。

与讨论 Levinson 问题时相同，假设(标量)观测序列是广义平稳的，即对于所有整数 n 和 l，$C_Y(n,l)\triangleq\mathrm{cov}(Y_n,Y_l)=C_Y(n-l,0)=C_Y(n-l)$。同时，假设观测序列和(标量)序列$\{X_n\}_{n=-\infty}^{\infty}$是联合平稳的，即对于所有整数 n 和 l，$C_{XY}(t,n)\triangleq\mathrm{cov}(X_t,Y_n)=C_{XY}(t-n,0)\equiv C_{XY}(t-n)$(为不失一般性，仍假设 X_n 和 Y_n 的均值为零)。此外，假设观测样本的个数是无限的，并讨论以下两种滤波问题：非因果 Wiener-Kolmogorov 滤波问题，其中 $a=-\infty$，$b=+\infty$；因果 Wiener-Kolmogorov 滤波问题，其中 $a=-\infty$，$b=t$。首先讨论求解过程相对简单的非因果滤波问题。

5.4.1 非因果 Wiener-Kolmogorov 滤波

在这一类滤波问题中，基于全部时刻的观测样本来求解 t 时刻的估计值，其中 $-\infty<t<\infty$。因此，称其为非因果 Wiener-Kolmogorov 问题，从而有

$$\hat{X}_t=\sum_{n=-\infty}^{\infty}h_{t,n}Y_n \tag{5.4.1}$$

⊖ 实际上，该结论可由 Kalman-Bucy 滤波器的高斯分布和线性性质推导出。上述模型包含了高斯模型，因此如果对于上述模型存在某个最优线性滤波器结构，那么 Kalman-Bucy 将不再是高斯分布这一特殊情况下的全局最优滤波器。

如果将上式看作对序列$\{Y_n\}_{n=-\infty}^{\infty}$的线性滤波操作，则相应的估计问题不受因果条件的约束，即冲激响应$\{h_{t,n}\}_{n=-\infty}^{\infty}$可以不满足$h_{t,n}=0$，$t>n$。这意味着，对"当前"时刻$t$的估计可能会取决于未来时刻$n>t$。显然，对于实时估计问题，应该只关注因果滤波器。然而，对于可事先存储的非实时数据应用或者t为空间参数而不是时间参数的情况(如图像或阵列处理)，因果性并非必要条件。

上述问题的 Wiener-Hopf 方程(5.3.18)为

$$C_{XY}(t,l)=\sum_{n=-\infty}^{\infty}h_{t,n}C_Y(n,l),\quad -\infty<l<\infty \tag{5.4.2}$$

由平稳假设可得

$$C_{XY}(t-l)=\sum_{n=-\infty}^{\infty}h_{t,n}C_Y(n-l),\quad -\infty<l<\infty \tag{5.4.3}$$

为了便于分析，定义新变量$\tau=t-l$，进而可将上式写为

$$C_{XY}(\tau)=\sum_{n=-\infty}^{\infty}h_{t,n}C_Y(n+\tau-t),\quad -\infty<\tau<\infty \tag{5.4.4}$$

再进行变量代换$\alpha=t-n$，上述 Wiener-Hopf 方程变为

$$C_{XY}(\tau)=\sum_{\alpha=-\infty}^{\infty}h_{t,t-\alpha}C_Y(\alpha-\tau),\quad -\infty<\tau<\infty \tag{5.4.5}$$

上式中，变量t仅出现在系数序列$\{h_{t,t-\alpha}\}_{\alpha=-\infty}^{\infty}$中，这表明如果上述 Wiener-Hopf 方程有解，则可得到与t无关的解。即最优解$\{h_{t,n}\}_{n=-\infty}^{\infty}$可选为$h_{t,t-\alpha}$仅与$\alpha$有关的形式，或者等价地写为$h_{t,t-\alpha}=h_{\alpha,0}$(对所有整数$t$和$\alpha$均成立)。因此，如果上式有解，则可以采用满足$h_{t,n}=h_{t-n,0}\triangleq h_{t-n}$的时不变(或平移不变)形式，其中，为方便起见，省略了第二个下标。根据该观测模型，并利用$C_Y(\alpha-\tau)=C_Y(\tau-\alpha)$，Wiener-Hopf 方程可写为

$$C_{XY}(\tau)=\sum_{\alpha=-\infty}^{\infty}h_\alpha C_Y(\tau-\alpha),\quad -\infty<\tau<\infty \tag{5.4.6}$$

上式右边为序列$\{h_n\}_{n=-\infty}^{\infty}$和$\{C_Y(n)\}_{n=-\infty}^{\infty}$的离散时间卷积。因此，式(5.4.6)实际上是一个卷积方程，且可以被转换为频域的简单代数方程。特别地，假设下列离散时间傅里叶变换存在

$$H(\omega)\triangleq\sum_{n=-\infty}^{\infty}h_n e^{-i\omega n},\quad -\pi\leqslant\omega\leqslant\pi \tag{5.4.7}$$

$$\phi_{XY}(\omega)\triangleq\sum_{n=-\infty}^{\infty}C_{XY}(n)e^{-i\omega n},\quad -\pi\leqslant\omega\leqslant\pi \tag{5.4.8}$$

及

$$\phi_Y(\omega)\triangleq\sum_{n=-\infty}^{\infty}C_Y(n)e^{-i\omega n},\quad -\pi\leqslant\omega\leqslant\pi \tag{5.4.9}$$

则上述 Wiener-Hopf 方程变为

$$\phi_{XY}(\omega)=H(\omega)\phi_Y(\omega),\quad -\pi\leqslant\omega\leqslant\pi \tag{5.4.10}$$

其中，H是滤波器$\{h_n\}_{n=-\infty}^{\infty}$对应的传递函数，$\phi_Y$是序列$\{Y_n\}_{n=-\infty}^{\infty}$的功率谱密度(或频谱)，$\phi_{XY}$是序列$\{X_n\}_{n=-\infty}^{\infty}$和$\{Y_n\}_{n=-\infty}^{\infty}$的互功率谱密度。

利用式(5.4.10)，求解 Wiener-Hopf 方程易得最优估计所对应的传递函数，即

$$H(\omega)=\frac{\phi_{XY}(\omega)}{\phi_{Y}(\omega)},\quad -\pi\leqslant\omega\leqslant\pi \tag{5.4.11}$$

进而可得相应(时不变)滤波器系数为[⊖]

$$h_n=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\phi_{XY}(\omega)}{\phi_{Y}(\omega)}\mathrm{e}^{\mathrm{i}\omega n}\mathrm{d}\omega,\quad n\in\mathbf{Z} \tag{5.4.12}$$

基于前述假设，式(5.4.11)或式(5.4.12)给出了非因果 Wiener-Kolmogorov 问题中的最优线性估计。在采用具体示例对上述结果进行说明之前，有必要先对采用上述滤波器时所导致的均方误差进行讨论。定义均方误差最小值表达式为

$$\mathrm{MMSE}\triangleq\min_{\hat{X}_t\in H_{-\infty}^{\infty}}E\{(X_t-\hat{X})^2\}$$

即

$$\mathrm{MMSE}=E\{(X_t-\hat{X}_t)^2\}$$

其中，$\hat{X}_t$ 是最优估计值。可将上式进一步推导为

$$\begin{aligned}\mathrm{MMSE}&=E\{(X_t-\hat{X}_t)^2\}\\&=E\{(X_t-\hat{X}_t)X_t\}-E\{(X_t-\hat{X}_t)\hat{X}_t\}\\&=E\{X_t^2\}-E\{\hat{X}_tX_t\}\end{aligned} \tag{5.4.13}$$

其中，由正交性原理可知 $E\{(X_t-\hat{X}_t)\hat{X}_t\}=0$。

首先分析式(5.4.13)右边的第二项，可得

$$\begin{aligned}E\{\hat{X}_tX_t\}&=E\left\{\left(\sum_{n=-\infty}^{\infty}h_{t-n}Y_n\right)X_t\right\}=\sum_{n=-\infty}^{\infty}h_{t-n}E\{Y_nX_t\}\\&=\sum_{n=-\infty}^{\infty}h_{t-n}C_{XY}(t-n)=\sum_{\alpha=-\infty}^{\infty}h_{\alpha}C_{XY}(\alpha)\end{aligned} \tag{5.4.14}$$

其中，定义下标 $\alpha=t-n$。注意，式(5.4.14)右边是 $\{h_n\}_{n=-\infty}^{\infty}$ 与 $\{C_{XY}(-n)\}_{n=-\infty}^{\infty}$ 卷积的第 0 项，从而有

$$E\{\hat{X}_tX_t\}=\frac{1}{2\pi}\int_{-\pi}^{\pi}H(\omega)\overline{\phi}_{XY}(\omega)\mathrm{d}\omega \tag{5.4.15}$$

其中，$\overline{\phi}_{XY}(\omega)$ 是序列 $\{C_{XY}(-n)\}_{n=-\infty}^{\infty}$ 的离散时间傅里叶变换，即

$$\overline{\phi}_{XY}(\omega)=\sum_{n=-\infty}^{\infty}C_{XY}(-n)\mathrm{e}^{-\mathrm{i}\omega n},\quad -\pi\leqslant\omega\leqslant\pi \tag{5.4.16}$$

令式(5.4.16)中 $\alpha=-n$，有

$$\overline{\phi}_{XY}(\omega)=\sum_{\alpha=-\infty}^{\infty}C_{XY}(\alpha)\mathrm{e}^{\mathrm{i}\omega\alpha}=\phi_{XY}^{*}(\omega) \tag{5.4.17}$$

其中，上标“*”表示复共轭。由式(5.4.11)和式(5.4.15)可得

$$E\{\hat{X}_tX_t\}=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\phi_{XY}(\omega)}{\phi_{Y}(\omega)}\phi_{XY}^{*}(\omega)\mathrm{d}\omega=\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{|\phi_{XY}(\omega)|^2}{\phi_{Y}(\omega)}\mathrm{d}\omega \tag{5.4.18}$$

假设序列 $\{X_n\}_{n=-\infty}^{\infty}$ 是广义平稳的，其频谱为 ϕ_X，则可以将式(5.4.13)右边的第一项写为

⊖ 此处及本书其他各处，**Z** 均表示整数集合。

$$E\{X_t^2\} = C_X(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi_X(\omega)\mathrm{d}\omega$$

由上式可得

$$\mathrm{MMSE} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\phi_X(\omega) - \frac{|\phi_{XY}(\omega)|^2}{\phi_Y(\omega)}\right]\mathrm{d}\omega \tag{5.4.19}$$

上式可改写为

$$\mathrm{MMSE} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[1 - \frac{|\phi_{XY}(\omega)|^2}{\phi_X(\omega)\phi_Y(\omega)}\right]\phi_X(\omega)\mathrm{d}\omega \tag{5.4.20}$$

因此，最优滤波器的性能可由 ϕ_X 加权下的函数$[1-|\phi_{XY}|^2/\phi_X\phi_Y]$的积分得到。由互功率谱的性质可得，$|\phi_{XY}(\omega)|^2 \leqslant \phi_X(\omega)\phi_Y(\omega)$，$-\pi\leqslant\omega\leqslant\pi$，且对所有的 $\omega\in[-\pi,\pi]$，当且仅当序列$\{X_n\}_{n=-\infty}^{\infty}$和$\{Y_n\}_{n=-\infty}^{\infty}$完全相关时，等号成立(即 $X_t\in H_{-\infty}^{\infty}$ 对所有 $t\in\mathbf{Z}$ 成立)。因此，式(5.4.20)表明随着序列$\{X_n\}_{n=-\infty}^{\infty}$和$\{Y_n\}_{n=-\infty}^{\infty}$由完全无关$[\phi_{XY}(\omega)=0,\ -\pi\leqslant\omega\leqslant\pi]$到完全相关，最小均方误差的取值由 $E\{X_t^2\}$变化到 0。可见，在取值范围的两个端点分别有$\hat{X}_t=E\{X_t\}\equiv 0$ 和$\hat{X}_t=X_t$。显然，需要关注的是取值范围中两个端点之间的部分，即$0<\mathrm{MMSE}<E\{X_t^2\}$的部分。

为了对非因果 Wiener-Kolmogorov 滤波进行举例说明，给出如下示例。

例 5.4.1 加性噪声中的信号估计

考虑如下观测模型

$$Y_n = S_n + N_n,\quad n\in\mathbf{Z} \tag{5.4.21}$$

其中，序列$\{S\}_{n=-\infty}^{\infty}$和$\{N_n\}_{n=-\infty}^{\infty}$分别表示不相关、零均值且广义平稳的信号和噪声。假设在 t 时刻期望观测到的量为$t+\lambda$ 时刻的信号，且 λ 为整数，即

$$X_t = S_{t+\lambda} \tag{5.4.22}$$

上述问题在 $\lambda=0$、$\lambda>0$ 和 $\lambda<0$ 时分别代表对信号的滤波、预测和平滑。分别用 ϕ_S 和 ϕ_N 代表信号和噪声的功率谱，易得

$$\phi_Y(\omega) = \phi_S(\omega) + \phi_N(\omega),\quad -\pi\leqslant\omega\leqslant\pi \tag{5.4.23}$$

$$\phi_{XY}(\omega) = \mathrm{e}^{\mathrm{i}\omega\lambda}\phi_S(\omega),\quad -\pi\leqslant\omega\leqslant\pi \tag{5.4.24}$$

和

$$\phi_X(\omega) = \phi_S(\omega),\quad -\pi\leqslant\omega\leqslant\pi \tag{5.4.25}$$

由式(5.4.11)、式(5.4.23)和式(5.4.24)可得，最优非因果滤波器的传递函数为

$$H(\omega) = \frac{\mathrm{e}^{\mathrm{i}\omega\lambda}\phi_S(\omega)}{\phi_S(\omega)+\phi_N(\omega)},\quad -\pi\leqslant\omega\leqslant\pi \tag{5.4.26}$$

上述滤波器的解释非常直观：$\mathrm{e}^{\mathrm{i}\omega\lambda}$ 是具有单位幅度的相位项，在时域对应 λ 个单位时延。因此，这一项仅对数据序列进行时移操作，与期望对 $t+\lambda$ 时刻的信号进行估计一致。上述时延在 $\lambda\leqslant 0$ 和 $\lambda>0$ 时，分别对应因果和非因果两种情况。滤波器传递函数中的另一项是它的幅度项 $\phi_S/(\phi_S+\phi_N)$，代表滤波器的增益，可将这一项改写为

$$|H(\omega)| = \frac{\phi_S(\omega)/\phi_N(\omega)}{\phi_S(\omega)/\phi_N(\omega)+1},\quad -\pi\leqslant\omega\leqslant\pi \tag{5.4.27}$$

易知，上式在 $\phi_S(\omega)/\phi_N(\omega)$取 0 到无穷大时由 0 变化到 1(注意，功率谱是实值非负函数)。

事实上，如果 $\phi_S(\omega)/\phi_N(\omega) \ll 1$，则 $|H(\omega)| \approx 0$；如果 $\phi_S(\omega)/\phi_N(\omega) \gg 1$，则 $|H(\omega)| \approx 1$。可以将 $\phi_S(\omega)/\phi_N(\omega)$ 看作在频率 ω 处对信噪比(信噪功率比)的一种度量。可见，如果某个频率处的噪声分量占优，则相应的滤波器增益趋近于 0；如果信号分量占优，则滤波器增益趋近于 1。在上述两个边界取值之间，增益的选取使得信号失真(由小于 1 的增益造成)和噪声残留(由大于 0 的增益导致)达到一种平衡。

由式(5.4.20)和式(5.4.23)～式(5.4.25)可得，非因果 Wiener-Kolmogorov 滤波器的性能可由下式给出(易知时延 λ 与滤波器性能无关)

$$\mathrm{MMSE} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\phi_S(\omega)\phi_N(\omega)}{\phi_S(\omega)+\phi_N(\omega)}\mathrm{d}\omega \tag{5.4.28}$$

由于 $\phi_S(\omega)/[\phi_S(\omega)+\phi_N(\omega)] \leqslant 1$ 且 $\phi_N(\omega)/[\phi_S(\omega)+\phi_N(\omega)] \leqslant 1$，因此，式(5.4.28)表示最小均方误差不会大于信号平均功率 $\frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_S(\omega)\mathrm{d}\omega$ 或者噪声平均功率 $\frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_N(\omega)\mathrm{d}\omega$，即

$$\mathrm{MMSE} \leqslant \min\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_S(\omega)\mathrm{d}\omega, \frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_N(\omega)\mathrm{d}\omega\right\} \tag{5.4.29}$$

当 $\phi_S(\omega)/\phi_N(\omega)$ 在 $[-\pi, \pi]$ 内均趋近于 0 时，取上式两个极限值中的第一项，此时最优滤波器变为全阻滤波器 $[H(\omega) \approx 0]$；而当 $\phi_S(\omega)/\phi_N(\omega)$ 在 $[-\pi, \pi]$ 内均趋近于无穷时，取上式两个极限值中的第二项，此时最优滤波器为全通滤波器 $[H(\omega) \approx 1]$。式(5.4.28)同时表明，当且仅当 $\phi_S(\omega)\phi_N(\omega)=0$ 对所有 $\omega \in [-\pi, \pi]$ 均成立时，最小均方误差才为 0。比如，当信号与噪声分别占据不同频带时，就会出现这种情况。◀

5.4.2 因果 Wiener-Kolmogorov 滤波

如前所述，非因果线性估计不适用于需要进行实时估计的场合。由于许多应用中都需要进行实时估计，因此有必要对因果 Wiener-Kolmogorov 问题进行研究。在这类问题中，希望基于序列 $\{Y_n\}_{n=-\infty}^{\infty}$ 观测样本中 t 时刻之前的部分实现对 X_t 的估计，即对应 $a=-\infty$、$b=t$ 时的情况。

为了求解上述问题，首先，由于集合 $H_{-\infty}^{t}$ 是集合 $H_{-\infty}^{\infty}$ 的子集。因此，如果非因果 Wiener-Kolmogorov 问题的解刚好是因果的，那么它也能用于求解因果 Wiener-Kolmogorov 问题。遗憾的是，除特殊情况外，非因果问题的解一般是严格非因果的。然而，在因果和非因果 Wiener-Kolmogorov 问题的解之间存在某种确定关系。如果令 $\hat{X}_t$ 和 $\widetilde{X}_t$ 分别表示已知模型的因果和非因果 Wiener-Kolmogorov 估计，则有 $(X_t-\hat{X}_t)=(\widetilde{X}_t-\hat{X}_t)+(X_t+\widetilde{X}_t)$，且对任意的 $Z \in H_{-\infty}^{t}$，有

$$0 = E\{(X_t-\hat{X}_t)Z\} = E\{(\widetilde{X}_t-\hat{X}_t)Z\} + E\{(X_t-\widetilde{X}_t)Z\} \tag{5.4.30}$$

进一步地，将正交性原理应用于 $\hat{X}_t$ 并利用 $Z \in H_{-\infty}^{\infty}$ (由于 $H_{-\infty}^{t} \subset H_{-\infty}^{\infty}$)，则有 $E\{(X_t-\widetilde{X}_t)Z\}$ 项为 0。因此，由式(5.4.30)可得

$$E\{(\widetilde{X}_t-\hat{X}_t)Z\} = 0, \quad Z \in H_{-\infty}^{t} \tag{5.4.31}$$

由式(5.4.31)和正交性原理可得，$\hat{X}_t$ 是 $H_{-\infty}^{t}$ 中所有估计值 $\widetilde{X}_t$ 的最小均方估计。由于 $\hat{X}_t$ 是 X_t 向 $H_{-\infty}^{\infty}$ 的投影，$\hat{X}_t$ 可以由下列步骤得到：首先将 X_t 向 $H_{-\infty}^{\infty}$ 投影得到 $\widetilde{X}_t$，然后将 $\widetilde{X}_t$ 向 $H_{-\infty}^{t}$ 进行投影。以空间 $\mathbf{R}^3$ 下的标准正交坐标系为例，其中，三个坐标轴分别以 x、

y 和 z 表示，则上述结论的几何解释为：为了得到 $\mathbf{R}^3$ 中某向量向 X 轴的投影，可以先将该向量向 x-y 平面(或 x-z 平面)投影，然后再将投影结果向 x 轴投影。由于 x 轴是 x-y 平面(或 x-z 平面)的子集，因此上述操作显然是可行的。

基于上述讨论，可以将非因果最小均方误差估计结果作为求解因果最小均方误差估计的出发点

$$\widetilde{X}_t = \sum_{n=-\infty}^{\infty} \widetilde{h}_{t-n} Y_n \tag{5.4.32}$$

其中，$\{\widetilde{h}_n\}_{n=-\infty}^{\infty}$ 由式(5.4.12)给出。为了实现 $\widetilde{X}_t$ 向 $\{Y_n\}_{n=-\infty}^{t}$ 构成的线性估计集合的投影，可以对式(5.4.32)进行一种简单的截断操作，即

$$\overline{X}_t \triangleq \sum_{n=-\infty}^{t} \widetilde{h}_{t-n} Y_n \tag{5.4.33}$$

如果 $\overline{X}_t$ 是 $\widetilde{X}_t$ 向 $H_{-\infty}^t$ 的投影，则对于所有的 $m \leqslant t$，误差 $\widetilde{X}_t - \overline{X}_t = \sum_{n=t+1}^{\infty} \widetilde{h}_{t-n} Y_n$ 必与 Y_m 正交。然而，由于 $\{Y_n\}_{n=-\infty}^{\infty}$ 和 $\{Y_n\}_{n=-\infty}^{t}$ 往往是相关的，且系数序列 $\{\widetilde{h}_n\}_{n=-\infty}^{\infty}$ 满足不同的正交性条件，这使得通常情况下，$(\widetilde{X}_t - \overline{X}_t)$ 不会与 $H_{-\infty}^t$ 正交。但是，当 $\{Y_n\}_{n=-\infty}^{\infty}$ 由一组非相关随机变量构成时，式(5.4.33)中的 $\overline{X}_t$ 将会符合所需的正交性条件，这时有

$$E\{(\widetilde{X}_t - \overline{X}_t)Y_m\} = \sum_{n=t+1}^{\infty} \widetilde{h}_{t-n} E\{Y_n Y_m\} = \sigma^2 \sum_{n=t+1}^{\infty} \widetilde{h}_{t-n} \delta_{n,m} = 0, m \leqslant t \tag{5.4.34}$$

其中，$\delta_{n,m}$ 为 Kronecker 函数(当 $n=m$ 时，$\delta_{n,m}=1$；当 $n \neq m$ 时，$\delta_{n,m}=0$)，且 $\sigma^2 = E\{Y_n^2\}$ (假设 $\{Y_n\}_{n=-\infty}^{\infty}$ 均值为零且广义平稳)。

因此，由上述分析可知，如果首先利用因果线性算子将 $\{Y_n\}_{n=-\infty}^{\infty}$ 转化为等价的广义平稳序列 $\{Z_n\}_{n=-\infty}^{\infty}$ ($\{Z_n\}_{n=-\infty}^{\infty}$ 由非相关随机变量构成)，然后求解 $\{Z_n\}_{n=-\infty}^{\infty}$ 对 X_t 的最优非因果估计，则相应的因果 Wiener-Kolmogorov 估计可由上述非因果估计的直接截断得到[如式(5.4.33)所示]。既然序列 $\{Z_n\}_{n=-\infty}^{\infty}$ 应该具有恒定谱，则相应的因果估计值为

$$\hat{X}_t = \sum_{n=-\infty}^{t} \hat{h}_{t-n} Z_n \tag{5.4.35}$$

其中

$$\hat{h}_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_{XZ}(\omega) \mathrm{e}^{\mathrm{i}\omega n} \mathrm{d}\omega = C_{XZ}(n), \quad n \geqslant 0 \tag{5.4.36}$$

ϕ_{XZ} 和 C_{XZ} 分别为 $\{X_n\}_{n=-\infty}^{\infty}$ 和 $\{Z_n\}_{n=-\infty}^{\infty}$ 的互功率谱和互协方差。

与 3.2 节和 5.2 节中针对高斯检测估计问题提出的概念类似，将 $\{Y_n\}_{n=-\infty}^{\infty}$ 因果转换为等价的白序列[⊖] $\{Z_n\}_{n=-\infty}^{\infty}$ 是合理的。利用之前的分析结果，假设 $\hat{Y}_{t|t-1}$ 是基于 $\{Y_n\}_{n=-\infty}^{t-1}$ 对 Y_t 的最优线性预测，即假设对于所有 $Z \in H_{-\infty}^t$，$\hat{Y}_{t|t-1}$ 使 $E\{(Y_t - Z)^2\}$ 最小。同时，令 σ_t^2 为该预测的均方误差，即

$$\sigma_t^2 = E\{(Y_t - \hat{Y}_{t|t-1})^2\} \tag{5.4.37}$$

定义序列 $\{Z_n\}_{n=-\infty}^{\infty}$ 为

⊖ “白”指由零均值非相关随机变量构成的广义平稳序列，相应序列在 $\omega \in [-\pi, \pi]$ 上具有恒定功率谱。因此，与白光中不同波长的所有可见光分量均相等的情况类似，称其为“白序列”。

$$Z_n = \frac{Y_n - \hat{Y}_{n|n-1}}{\sigma_n}, \quad n \in \mathbf{Z} \tag{5.4.38}$$

且 $Z_n \in H_{-\infty}^n$。可得 $E\{Z_n^2\}=1$，$E\{Z_n\}=0$ 且 $\mathrm{cov}(Z_n, Z_m)=E\{Z_n Z_m\}$。当 $m<n$ 时，根据 $Z_m \in H_{-\infty}^{n-1}$ 时的正交性原理，有

$$E\{Z_n Z_m\} = \frac{1}{\sigma_n} E\{(Y_n - \hat{Y}_{n|n-1}) Z_m\} = 0$$

类似地，当 $m>n$ 时有 $E\{Z_n Z_m\}=0$ 成立。因此，白序列 $\{Z_n\}_{n=-\infty}^{\infty}$ 可由 $\{Y_n\}_{n=-\infty}^{\infty}$ 的因果线性变换得到。

如果能够证明利用 $\{Z_n\}_{n=-\infty}^{\infty}$ 和 $\{Y_n\}_{n=-\infty}^{\infty}$ 进行线性最小均方误差估计等价，则因果 Wiener-Kolmogorov 估计问题就可简化为单步线性预测问题。这是因为单步线性预测可用来对式(5.4.38)所得的观测样本进行预白化处理，利用预白化后的数据，根据式(5.4.35)和式(5.4.36)可直接得到 X_t 的因果估计。事实上，在 $\{Y_n\}_{n=-\infty}^{\infty}$ 的功率谱满足一定条件时，$\{Z_n\}_{n=-\infty}^{\infty}$ 与 $\{Y_n\}_{n=-\infty}^{\infty}$ 等价是成立的。为了说明这一点，现基于线性预测问题给出进一步分析。

1. 线性预测

考虑 $X_t = Y_{t+\lambda}$ 时的因果 Wiener-Kolmogorov 问题，其中，λ 为正整数。注意，当 $\lambda=1$ 时，该问题与 Levinson 问题类似，只是这里的观测样本在时域上可回溯至 $-\infty$ 时刻。为了对该问题进行求解，首先给出线性预测理论中的核心结论。

命题 5.4.1　谱分解定理

假设 $\{Y_n\}_{n=-\infty}^{\infty}$ 的功率谱服从所谓的 Paley-Wiener 条件，即

$$c_0 \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\phi_Y(\omega)\,\mathrm{d}\omega > -\infty \tag{5.4.39}$$

则 ϕ_Y 可以写作 $\phi_Y(\omega)=\phi_Y^+(\omega)\phi_Y^-(\omega)$，$-\pi\leqslant\omega\leqslant\pi$，其中，$\phi_Y^+$ 和 ϕ_Y^- 是满足 $|\phi_Y^+(\omega)|^2 = |\phi_Y^-(\omega)|^2=\phi_Y(\omega)$ 的两个函数，即

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \phi_Y^+(\omega)\mathrm{e}^{\mathrm{i}n\omega}\,\mathrm{d}\omega = 0, \quad n<0 \tag{5.4.40a}$$

且

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \phi_Y^-(\omega)\mathrm{e}^{\mathrm{i}n\omega}\,\mathrm{d}\omega = 0, \quad n>0 \tag{5.4.40b}$$

此外，当 ϕ_Y^+ 和 ϕ_Y^- 被分别替换为 $1/\phi_Y^+$ 和 $1/\phi_Y^-$ 时，式(5.4.40a)和式(5.4.40b)依然成立[⊖]。

证明　完整的证明过程由 Ash 和 Gardner(1975)给出。在这里简要给出证明过程所涉及的主要思想。

首先，由于 $\log(x)$是 x 的凹函数，则由 Jansen 不等式可得

$$c_0 \leqslant \log\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_Y(\omega)\,\mathrm{d}\omega\right) = \log(E\{Y^2\}) < \infty$$

因此，式(5.4.39)中所述条件与条件 $|c_0|<\infty$ 等价。从而，可以给出 $\log\phi_Y$ 的离散傅里叶变换形式

⊖ 注意，式(5.4.39)表明 ϕ_Y 是非零的。因为 $|\phi_Y^+(\omega)|^2=|\phi_Y^-(\omega)|^2=\phi_Y(\omega)$，可得 ϕ_Y^+ 和 ϕ_Y^- 也必为非零的，从而 $1/\phi_Y^+$ 和 $1/\phi_Y^-$ 均是有意义的。当 $1/\phi_Y^+$ 满足式(5.4.40a)时，称 ϕ_Y^+ 为最小相位。

$$\log\phi_Y(\omega) = \sum_{n=-\infty}^{\infty} c_n \mathrm{e}^{-\mathrm{i}\omega n}, \quad -\pi < \omega \leqslant \pi \tag{5.4.41a}$$

其中

$$c_n \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}n\omega} \log\phi_Y(\omega), \quad n \in \mathbf{Z} \tag{5.4.41b}$$

由式(5.4.41a)可得

$$\phi_Y(\omega) = \exp\left\{\sum_{n=-\infty}^{\infty} c_n \mathrm{e}^{-\mathrm{i}\omega n}\right\} = \phi_Y^+(\omega)\phi_Y^-(\omega), \quad -\pi < \omega \leqslant \pi \tag{5.4.42}$$

其中

$$\phi_Y^+(\omega) \triangleq \exp\left\{\frac{c_0}{2} + \sum_{n=1}^{\infty} c_n \mathrm{e}^{-\mathrm{i}\omega n}\right\} \tag{5.4.43}$$

且

$$\phi_Y^-(\omega) \triangleq \exp\left\{\frac{c_0}{2} + \sum_{n=-\infty}^{-1} c_n \mathrm{e}^{-\mathrm{i}\omega n}\right\} \tag{5.4.44}$$

接下来，证明 ϕ_Y^+ 和 ϕ_Y^- 满足谱分解定理所需的性质。为此，首先需要指出的是所有功率谱密度均是偶对称函数[即 $\phi_Y(-\omega)=\phi_Y(\omega)$]，该结论可直接由 C_Y 的偶对称性得到。由 $\phi_Y(\omega)$ 的偶对称性可得 $\log\phi_Y(\omega)$ 也是偶对称的，从而进一步可得式(5.4.41b)中的系数 $\{c_n\}_{n=-\infty}^{\infty}$ 均为实数且是偶对称的(即 $c_{-n}=c_n=c_n^*$)。从而有

$$\phi_Y^-(\omega) = \exp\left\{\frac{c_0}{2} + \sum_{n=1}^{\infty} c_{-n} \mathrm{e}^{\mathrm{i}\omega n}\right\} = \exp\left\{\frac{c_0}{2} + \sum_{n=1}^{\infty} c_n \mathrm{e}^{\mathrm{i}\omega n}\right\} = [\phi^+(\omega)]^* \tag{5.4.45}$$

式(5.4.45)表明

$$|\phi_Y^-(\omega)|^2 = |\phi_Y^+(\omega)|^2 = \phi_Y^+(\omega)[\phi_Y^+(\omega)]^* = \phi_Y^+(\omega)\phi_Y^-(\omega) = \phi_Y(\omega)$$

为了证明式(5.4.40a)，首先考虑到 e^z 具有如下的幂级数展开式

$$\mathrm{e}^z = \sum_{k=0}^{\infty} \frac{z^k}{k!} \tag{5.4.46}$$

从而有

$$\phi_Y^+(\omega) = \mathrm{e}^{c_0/2}\left(\sum_{k=0}^{\infty}\left(\sum_{n=1}^{\infty} c_n \mathrm{e}^{-\mathrm{i}\omega n}\right)^k / k!\right) \tag{5.4.47}$$

考察式(5.4.47)可知，等式右边仅包含 $\mathrm{e}^{-\mathrm{i}\omega}$ 的非负次幂。这表明 $\phi_Y^+(\omega)$ 的负数下标的傅里叶分量均为0，即式(5.4.40a)。式(5.4.40b)及其在 $1/\phi_Y^+$ 和 $1/\phi_Y^-$ 下的类似情况也可通过相似证明过程加以论证。 □

值得注意的是，将 ϕ_Y 进行谱分解得到因果部分与反因果部分的乘积形式与将协方差矩阵进行 Cholesky 分解为上三角因子和下三角因子是类似的。

可以利用命题5.4.1中的谱分解去求解线性最小均方误差预测问题。为此，在接下来的讨论中，若没有特别说明，均假设 ϕ_Y 满足式(5.4.39)。考虑具有传递函数 $H(\omega)=1/\phi_Y^+(\omega)$ 的时不变线性滤波器。可见，这是将式(5.4.40a)用 $1/\phi_Y^+$ 代换所对应的因果滤波器。假设将序列 $\{Y_n\}_{n=-\infty}^{\infty}$ 作用于该滤波器，并设 $\{W_n\}_{n=-\infty}^{\infty}$ 为相应的输出序列。在二阶随机序列分析中，"广义平稳过程作用于线性时不变滤波器后的输出仍为广义平稳的，且输出过程的功率谱为 $|H(\omega)|^2\phi(\omega)$"这一结论是众所周知的。其中，$H$ 为滤波器传递函数而

ϕ 为输入功率谱[证明由 Wong(1983)给出]。因此，$\{W_n\}_{n=-\infty}^{\infty}$ 为广义平稳序列，对应的功率谱为

$$\phi_W(\omega)=\left|\frac{1}{\phi_Y^+(\omega)}\right|^2\phi_Y(\omega)=\frac{\phi_Y(\omega)}{|\phi_Y^+(\omega)|^2}=1,\quad -\pi\leqslant\omega\leqslant\pi \tag{5.4.48}$$

由于恒定功率谱对应白序列，可知滤波器 $1/\phi_Y^+(\omega)$ 是 $\{Y_n\}_{n=-\infty}^{\infty}$ 的白化滤波器。此外，$\{Y_n\}_{n=-\infty}^{\infty}$ 可由 $\{W_n\}_{n=-\infty}^{\infty}$ 作用于具有传递函数 $\phi_Y^+(\omega)$ 的因果滤波器得到，即

$$Y_t=\sum_{n=-\infty}^{t}f_{t-n}W_n,\quad t\in\mathbf{Z} \tag{5.4.49}$$

其中

$$f_n=\frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_Y^+(\omega)\mathrm{e}^{\mathrm{i}\omega n}\mathrm{d}\omega,\quad n\geqslant 0 \tag{5.4.50}$$

式(5.4.49)给出了输入为白序列时，线性时不变滤波器输出 $\{Y_n\}_{n=-\infty}^{\infty}$ 的表达式。该表达式可用于导出序列 $\{Y_n\}_{n=-\infty}^{\infty}$ 的最优线性预测。特别地，当 $\lambda>0$ 时

$$Y_{t+\lambda}=\sum_{n=-\infty}^{t+\lambda}f_{t+\lambda-n}W_n=\sum_{n=t+1}^{t+\lambda}f_{t+\lambda-n}W_n+\sum_{n=-\infty}^{t}f_{t+\lambda-n}W_n \tag{5.4.51}$$

由于 $\{W_n\}_{n=-\infty}^{\infty}$ 是白的，变量 $W_{t+1},\cdots,W_{t+\lambda}$ 与 $\{W_n\}_{n=-\infty}^{t}$ 正交，同时，由式(5.4.49)可知它们与 $H_{-\infty}^t$ 正交。整理后可得

$$Y_{t+\lambda}-\sum_{n=-\infty}^{t}f_{t+\lambda-n}W_n=\sum_{n=t+1}^{t+\lambda}f_{t+\lambda-n}W_n$$

从而可得 $Y_{t+\lambda}-\sum_{n=-\infty}^{t}f_{t+\lambda-n}W_n$ 与 $H_{-\infty}^t$ 正交。由于 $\sum_{n=-\infty}^{t}f_{t+\lambda-n}W_n\in H_{-\infty}^t$，则由正交性原理可知，基于 $\{Y_n\}_{n=-\infty}^{t}$ 对 $Y_{t+\lambda}$ 的最优线性预测为

$$\hat{Y}_{t+\lambda}=\sum_{n=-\infty}^{t}f_{t+\lambda-n}W_n \tag{5.4.52}$$

式(5.4.52)对应的线性预测滤波器可以理解为两个线性时不变滤波器的级联，如图 5.4.1所示。其中，第一个滤波器是传递函数为 $1/\phi_Y^+(\omega)$ 的白化滤波器。第二个滤波器具有如下冲激响应

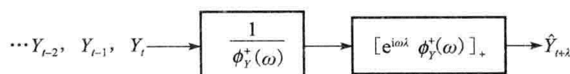


图 5.4.1 最优纯预测滤波器

$$\begin{cases}f_{n+\lambda}, & n\geqslant 0\\ 0, & n<0\end{cases} \tag{5.4.53}$$

如果定义算子 $[H(\omega)]_+$ 如下

$$[H(\omega)]_+=\sum_{n=0}^{\infty}h_n\mathrm{e}^{-\mathrm{i}\omega n} \tag{5.4.54}$$

其中，$h_n=\frac{1}{2\pi}\int_{-\pi}^{\pi}H(\omega)\mathrm{e}^{\mathrm{i}\omega n}\mathrm{d}\omega$，则式(5.4.53)对应的滤波器具有传递函数 $[\mathrm{e}^{\mathrm{i}\omega\lambda}\phi_Y^+(\omega)]_+$。因此，$\lambda$ 步最优线性预测的传递函数为

$$\frac{1}{\phi_Y^+(\omega)}[\mathrm{e}^{\mathrm{i}\omega\lambda}\phi_Y^+(\omega)]_+\triangleq H_\lambda(\omega) \tag{5.4.55}$$

综合式(5.4.49)和式(5.4.52)可得到最优线性预测的均方误差表达式。特别地，由

式(5.4.51)可得

$$\begin{aligned}\text{MMSE} &= \min_{Z\in H_{-\infty}^{t}} E\{(Y_{t+\lambda}-Z)^2\} = E\{(Y_{t+\lambda}-\hat{Y}_{t+\lambda})^2\} \\ &= E\left\{\left(\sum_{n=t+1}^{t+\lambda} f_{t+\lambda-n}W_n\right)^2\right\} = \sum_{n=t+1}^{t+\lambda} f_{t+\lambda-n}^2\end{aligned} \tag{5.4.56}$$

其中，最后一个等式可由 W_{t+1}，W_{t+2}，…与 $W_{t+\lambda}$ 之间的正交性得到。对式(5.4.56)中的变量下标进行变换后，可得

$$\text{MMSE} = \sum_{n=0}^{\lambda-1} f_n^2 \tag{5.4.57}$$

由于 f_n 是 ϕ_Y^+ 的 n 阶傅里叶系数，则系数 $\{f_n\}_{n=0}^{\lambda-1}$ 可由式(5.4.47)得到，尽管在 λ 较大时的求解过程会较为冗长。特别地，就一步预测($\lambda=1$)而言，由式(5.4.47)易知 $a_0=\mathrm{e}^{c_0/2}$。因此，一步预测对应的最小均方误差值为

$$\text{MMSE} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\phi_Y(\omega)\mathrm{d}\omega\right\} \tag{5.4.58}$$

该结果被称为 Kolmogorov-Szegö-Krein 公式。

式(5.4.51)和 $\{W_n\}_{n=-\infty}^{\infty}$ 的正交性表明[⊖]

$$\sum_{n=0}^{\infty} f_n^2 = E\{(Y_{t+\lambda})^2\} < \infty \tag{5.4.59}$$

对比式(5.4.57)和式(5.4.59)可知，如果期望预测得更远(即增大 λ)，则相应的最小均方预测误差将趋近于期望预测的量的均方值。也就是说，如果设定了想要预测 $\{Y_n\}_{n=-\infty}^{\infty}$ 的时刻(即设定 $t+\lambda$)并让 t 趋近于 $-\infty$，则当 $t\to-\infty$ 时的观测样本对 $Y_{t+\lambda}$ 的预测不再有用。即对于给定的 n，任意 Y_n 均无法由无限远的过去值得到。具有这一性质的序列被称为**纯非确定性的**，并且，当且仅当序列具有表达式 $\sum_{n=-\infty}^{t} f_{t-n}W_n$ 时才具有上述性质，其中，$\{W_n\}_{n=-\infty}^{\infty}$ 是白的且 $\sum_{n=0}^{\infty} f_n^2 < \infty$。命题 5.4.1 及后续分析表明 Paley-Wiener 条件是上述表达式成立的充分条件。同时，也可以证明，该条件也是上述表达式成立的必要条件[详见 Ash 和 Gardner(1975)的相关讨论]。

在明确了宽平稳序列的因果预白化理论之后，接下来开始讨论对广义因果 Wiener-Kolmogorov 问题的求解。特别地，假设 $\{Y_n\}_{n=-\infty}^{\infty}$ 满足 Paley-Wiener 条件，则可将其通过滤波器 $1/\phi_Y^+$ 后得到相应的白化结果(即白化为 $\{Z_n\}_{n=-\infty}^{\infty}$)。接下来，仅需要求解互功率谱 ϕ_{XZ}，并根据式(5.4.35)和式(5.4.36)求解级联于预白化滤波器之后的最优滤波器 $[\phi_{XZ}(\omega)]_+$ [符号 $[\cdot]_+$ 的定义与式(5.4.54)相同]。这个直观的操作表明，如果将 $\{Y_n\}_{n=-\infty}^{\infty}$ 通过传递函数为 H 的滤波器后得到 $\{Z_n\}_{n=-\infty}^{\infty}$，则有 $\phi_{XZ}=\phi_{XY}H^*$。因此，由 $H=1/\phi_Y^+$ 可得

⊖ 等价地，Parceval 公式给出

$$\sum_{n=0}^{\infty} f_n^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi}|\phi_Y^+(\omega)|^2\mathrm{d}\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_Y(\omega)\mathrm{d}\omega = E\{Y_t^2\},\quad t\in\mathbf{Z}$$

$$\phi_{XZ}(\omega)=\frac{\phi_{XY}(\omega)}{[\phi_Y^+(\omega)]^*}=\frac{\phi_{XY}(\omega)}{\phi_Y^-(\omega)} \tag{5.4.60}$$

则基于$\{Y_n\}_{n=-\infty}^{t}$估计 X_t 的因果 Wiener-Kolmogorov 滤波器具有如下传递函数

$$\frac{1}{\phi_Y^+(\omega)}\left[\frac{\phi_{XY}(\omega)}{\phi_Y^-(\omega)}\right]_+ \tag{5.4.61}$$

对比非因果 Wiener-Kolmogorov 滤波器的传递函数[详见式(5.4.11)]

$$\frac{\phi_{XY}(\omega)}{\phi_Y(\omega)}=\frac{1}{\phi_Y^+(\omega)}\left[\frac{\phi_{XY}(\omega)}{\phi_Y^-(\omega)}\right] \tag{5.4.62}$$

显然，上述两种滤波器均可以表示为一个因果预白化滤波器与另一个滤波器级联的形式。对于非因果滤波器，第二个滤波器为 ϕ_{XY}/ϕ_Y^-，而对于因果滤波器，第二个滤波器对应 ϕ_{XY}/ϕ_Y^- 的因果部分。

2. 有理谱的因式分解

设计因果 Wiener-Kolmogrov 滤波器的关键步骤是对观测谱 ϕ_Y 进行因式分解。由于可以将 ϕ_Y^+ 和 ϕ_Y^- 写为包含 $\log\phi_Y$ 的傅里叶系数$\{c_n\}_{n=-\infty}^{\infty}$的形式[详见式(5.4.43)及式(5.4.44)]，则上述因式分解可以通过对 c_n 的计算来实现(相应频谱的对数称为倒频谱)。然而，实际应用中，对所关心的大多数频谱来说，频谱的因式分解可以通过对复多项式的因式分解实现。下面主要讨论对这一类频谱的因式分解。

如果随机序列$\{Y_n\}_{n=-\infty}^{\infty}$的功率谱可以写为两个实三角多项式之比的形式，则称其为有理谱。如果 ϕ 为有理谱，则有

$$\phi_Y(\omega)=\frac{n_0+2\sum_{k=1}^{p}n_k\cos k\omega}{d_0+2\sum_{k=1}^{m}d_k\cos k\omega} \tag{5.4.63}$$

其中，m 和 p 均为正数且 n_0，…，n_p 和 d_0，…，d_m 为实数。实际应用中，当 m 和 p 足够大时，很多随机序列的功率谱均可近似为有理谱。由于有理谱必关于 $\omega=0$ 偶对称，式(5.4.62)中的多项式仅包含余弦项。

由于 $2\cos k\omega=e^{ik\omega}+e^{-ik\omega}$，式(5.4.63)中的频谱可写为

$$\phi_Y(\omega)=\frac{N(e^{i\omega})}{D(e^{i\omega})} \tag{5.4.64}$$

其中，N 和 D 均为复变量 z 的多项式，定义为

$$N(z)=\sum_{k=-p}^{p}n_{|k|}z^{-k} \tag{5.4.65a}$$

和

$$D(z)=\sum_{k=-m}^{m}d_{|k|}z^{-k} \tag{5.4.65b}$$

由于 $z^pN(z)$为 $2p$ 阶多项式，因此有 $2p$ 个根，且可以写为 $n_p\prod_{k=1}^{2p}(z-z_k)$。因此，可得

$$N(z)=n_pz^{-p}\prod_{k=1}^{2p}(z-z_k) \tag{5.4.66}$$

由于 $N(z)=N(1/z)$，则多项式的根 z_1，…，z_{2p}必互为倒数。即当根 z_k 存在时，必存在

另一个根 $1/z_k$。为表达方便，假设 z_1，…，z_{2p} 满足 $|z_1|\geqslant|z_2|\geqslant\cdots\geqslant|z_{2p}|$，则有

$$N(z)=B(z)B(1/z) \tag{5.4.67}$$

其中

$$B(z)=[(-1)^p n_p/z_1z_2\cdots z_p]^{1/2}\prod_{k=1}^{p}(z^{-1}-z_k) \tag{5.4.68}$$

由于 $|z_1|\geqslant|z_2|\geqslant\cdots\geqslant|z_{2p}|$，有 $|z_{2p}|=1/|z_1|$，$|z_{2p-1}|=1/|z_2|$，…，$|z_{p+1}|=1/|z_p|$ 成立，且满足 $|z_1|\geqslant|z_2|\geqslant\cdots\geqslant|z_p|\geqslant1$。进而可将 $B(z)$ 展开为

$$B(z)=\sum_{k=0}^{p}b_kz^{-k} \tag{5.4.69}$$

同理，多项式 $D(z)$ 可写为

$$D(z)=A(z)A(1/z) \tag{5.4.70}$$

其中

$$A(z)=[d_m/p_1p_2\cdots p_m]^{1/2}\prod_{k=1}^{m}(z^{-1}-p_k)=\sum_{k=0}^{m}a_kz^{-k} \tag{5.4.71}$$

其中，$|p_1|\geqslant|p_2|\geqslant\cdots\geqslant|p_m|\geqslant1$。

由上述分析可知，有理谱 $\phi_Y(\omega)$ 可写为

$$\phi_Y(\omega)=\frac{B(\mathrm{e}^{\mathrm{i}\omega})B(\mathrm{e}^{-\mathrm{i}\omega})}{A(\mathrm{e}^{\mathrm{i}\omega})A(\mathrm{e}^{-\mathrm{i}\omega})} \tag{5.4.72}$$

在后续分析中，假设 $B(z)$ 和 $A(z)$ 的根均不在单位圆 $|z|=1$ 上(即假设 $|z_p|>1$ 且 $|p_m|>1$)，这可保证 $\phi_Y(\omega)$ 是有上界的，且下界远离零值，从而满足 Paley-Wiener 条件]。易知 $B(\mathrm{e}^{\mathrm{i}\omega})/A(\mathrm{e}^{\mathrm{i}\omega})$ 和 $A(\mathrm{e}^{\mathrm{i}\omega})/B(\mathrm{e}^{\mathrm{i}\omega})$ 均为因果稳定的传递函数[详见 Oppenheim 和 Schafer (1975)的分析]，而 $B(\mathrm{e}^{-\mathrm{i}\omega})/A(\mathrm{e}^{-\mathrm{i}\omega})$ 和 $A(\mathrm{e}^{-\mathrm{i}\omega})/B(\mathrm{e}^{-\mathrm{i}\omega})$ 均为反因果稳定的传递函数。由上述分析和式(5.4.72)可得 ϕ_Y 的频谱因子为

$$\phi_Y^+(\omega)=B(\mathrm{e}^{\mathrm{i}\omega})/A(\mathrm{e}^{\mathrm{i}\omega}) \tag{5.4.73a}$$

和

$$\phi_Y^-(\omega)=[\phi_Y^+(\omega)]^*=B(\mathrm{e}^{-\mathrm{i}\omega})/A(\mathrm{e}^{-\mathrm{i}\omega}) \tag{5.4.73b}$$

$\{Y_n\}_{n=-\infty}^{\infty}$ 对应的白化滤波器由下式给出

$$\frac{1}{\phi_Y^+(\omega)}=\frac{A(\mathrm{e}^{\mathrm{i}\omega})}{B(\mathrm{e}^{\mathrm{i}\omega})} \tag{5.4.74}$$

同理，若用 $\{Z_n\}_{n=-\infty}^{\infty}$ 表示白化序列，则将 $\{Y_n\}_{n=-\infty}^{\infty}$ 作用于滤波器 $A(\mathrm{e}^{\mathrm{i}\omega})$ 时的输出与将 $\{Z_n\}_{n=-\infty}^{\infty}$ 作用于滤波器 $B(\mathrm{e}^{\mathrm{i}\omega})$ 时的输出相同。由式(5.4.69)和式(5.4.71)可知，$A(\mathrm{e}^{\mathrm{i}\omega})$ 和 $B(\mathrm{e}^{\mathrm{i}\omega})$ 的冲激响应分别为

$$\begin{cases}a_n, & 0\leqslant n\leqslant m\\ 0, & \text{其他}\end{cases} \tag{5.4.75}$$

和

$$\begin{cases}b_n, & 0\leqslant n\leqslant p\\ 0, & \text{其他}\end{cases} \tag{5.4.76}$$

这表明 $\{Y_n\}_{n=-\infty}^{\infty}$ 和 $\{Z_n\}_{n=-\infty}^{\infty}$ 具有如下关系

$$\sum_{k=0}^{m} a_k Y_{n-k} = \sum_{k=0}^{p} b_k Z_{n-k}, \quad n \in \mathbf{Z} \tag{5.4.77}$$

因此序列$\{Z_n\}_{n=-\infty}^{\infty}$满足下列递归公式

$$b_0 Z_n = -\sum_{k=1}^{p} b_k Z_{n-k} + \sum_{k=0}^{m} a_k Y_{n-k} \tag{5.4.78}$$

该递归公式给出一个有限维的线性数字滤波器，如图 5.4.2 所示(图中 z^{-1}表示单位延时)。

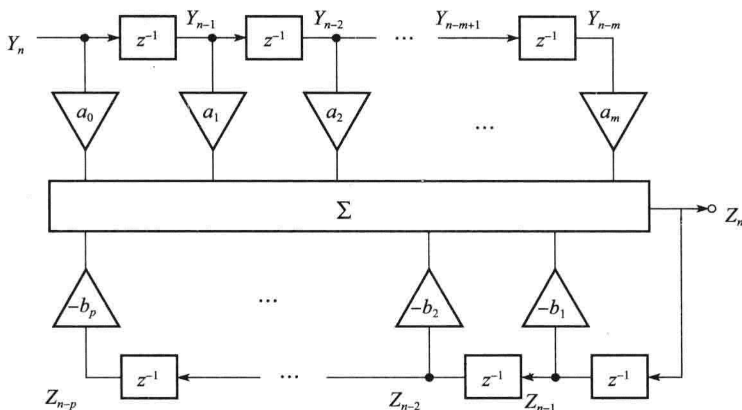


图 5.4.2 有理谱序列的白化滤波

根据式(5.4.77)，$\{Y_n\}_{n=-\infty}^{\infty}$可由$\{Z_n\}_{n=-\infty}^{\infty}$按照如下递归公式得到

$$a_0 Y_n = -\sum_{k=1}^{m} a_k Y_{n-k} + \sum_{k=0}^{p} b_k Z_{n-k}, \quad n \in \mathbf{Z} \tag{5.4.79}$$

由白序列按照上述方式产生的序列称为具有自回归阶数 m、滑动平均阶数 p 的自回归-滑动平均序列，或称为 ARMA(m, p)序列。ARMA(m, 0)序列也称为 m 阶自回归序列[AR(m)]，而 ARMA(0, p)序列也称为 p 阶滑动平均序列[MA(p)]。[当 $m=0$ 时，式(5.4.79)中的第一个求和项为 0。]ARMA 模型与 Kalman-Bucy 滤波中的状态空间模型密切相关，且其滤波性质可参考 Anderson 和 Moore(1979)的相关讨论。

例 5.4.2 广义马尔可夫序列的纯预测

广义马尔可夫模型是宽平稳随机序列分析中的一个简单但非常有用的相关结构模型

$$C_Y(n) = Pr^{|n|}, \quad n \in \mathbf{Z} \tag{5.4.80}$$

其中，$|r|<1$ 且 $P>0$。相应的功率谱为[详见 Thomas(1971)的推导]

$$\phi_Y(\omega) = \frac{P(1-r^2)}{1-2r\cos\omega + r^2} \tag{5.4.81}$$

显然式(5.4.81)为有理谱。利用 $2\cos\omega = e^{i\omega} + e^{-i\omega}$，有

$$\phi_Y(\omega) = \frac{P(1-r^2)}{1-re^{i\omega}-re^{-i\omega}+r^2} = \frac{P(1-r^2)}{(1-re^{-i\omega})(1-re^{i\omega})} = \frac{1}{A(e^{i\omega})A(e^{-i\omega})} \tag{5.4.82}$$

其中

$$A(z) = a_0 + a_1 z^{-1} \tag{5.4.83}$$

式(5.4.83)中，$a_0 = [P(1-r^2)]^{-1/2}$，$a_1 = -r[P(1-r^2)]^{-1/2}$。

假设需要对$\{Y_n\}_{n=-\infty}^{\infty}$进行 λ 步预测，则最优预测对应的传递函数可根据式(5.4.55)求

解。在本例中，传递函数可表示为

$$\hat{H}_\lambda(\omega) = A(e^{i\omega})\left[\frac{e^{i\omega\lambda}}{A(e^{i\omega})}\right]_+ \tag{5.4.84}$$

利用等比数列的相关结论：当$|x|<1$时，$\sum_{k=0}^{\infty} x^k = 1/(1-x)$。有下面的表达式成立

$$\frac{1}{A(z)} = \frac{1}{a_0(1-rz^{-1})} = \frac{1}{a_0}\sum_{n=0}^{\infty} r^n z^{-n} \tag{5.4.85}$$

由于$|r|<1$，上式在$|z|=1$时收敛，即$1/A(e^{i\omega}) = (1/a_0)\sum_{n=0}^{\infty} r^n e^{-i\omega n}$成立，继而可得到

$$\begin{aligned}\left[\frac{e^{i\omega\lambda}}{A(e^{i\omega})}\right]_+ &= \left[\frac{1}{a_0}\sum_{n=0}^{\infty} r^n e^{-i\omega(n-\lambda)}\right]_+ = \frac{1}{a_0}\sum_{n=\lambda}^{\infty} r^n e^{-i\omega(n-\lambda)} \\ &= \frac{1}{a_0}\sum_{l=0}^{\infty} r^{l+\lambda} e^{-i\omega l} = \frac{r^\lambda}{A(e^{i\omega})}\end{aligned} \tag{5.4.86}$$

综合式(5.4.84)和式(5.4.86)后求出最优预测的传递函数为$H_\lambda(\omega)=r^\lambda$，即该情况下的最优预测对应一个纯增益。显然，该预测器对应的冲激响应为$h_0=r^\lambda$且$h_n=0$，$n\neq 0$，进而

$$\hat{Y}_{t+\lambda} = r^\lambda Y_t \tag{5.4.87}$$

由式(5.4.87)和式(5.4.80)易得预测误差的均方值为

$$\begin{aligned}E\{(Y_{t+\lambda}-\hat{Y}_{t+\lambda})^2\} &= E\{Y_{t+\lambda}^2\} - E\{Y_{t+\lambda}\hat{Y}_{t+\lambda}\} \\ &= E\{Y_{t+\lambda}^2\} - r^\lambda E\{Y_{t+\lambda}Y_t\} = C_Y(0) - r^\lambda C_Y(\lambda) \\ &= P(1-r^{2\lambda})\end{aligned} \tag{5.4.88}$$

因为$|r|<1$，随着λ由 1 变化到∞，相应的预测误差由$(1-r^2)P$单调增加至P。◀

例 5.4.3 AR(m)序列的纯预测

由式(5.4.83)可知，AR(1)序列是一种广义的马尔可夫序列。特别地，按如下方式生成$\{Y_n\}_{n=-\infty}^{\infty}$

$$Y_{t+1} = rY_t + [P(1-r^2)]^{1/2} Z_{t+1}, \quad t\in\mathbf{Z} \tag{5.4.89}$$

其中，$\{Z_n\}_{n=-\infty}^{\infty}$是白的。因为$Z_{t+1}$与$\{Z_n\}_{n=-\infty}^{t}$和$\{Y_n\}_{n=-\infty}^{t}$均正交，则由式(5.4.89)可得$Y_{t+1}-rY_t$与$\{Y_n\}_{n=-\infty}^{t}$正交，且由正交性原理可得$rY_t$是基于$\{Y_n\}_{n=-\infty}^{t}$估计$Y_{t+1}$时的最小均方误差线性估计。这是式(5.4.87)中令$\lambda=1$时的情况。

类似地，对任意的自回归序列有

$$Y_{t+1} = -\sum_{k=1}^{m} a_k Y_{t+1-k} + b_0 Z_{t+1}, \quad t\in\mathbf{Z} \tag{5.4.90}$$

(为不失一般性，令$a_0=1$)，则$Y_{t+1}+\sum_{k=1}^{m} a_k Y_{t+1-k} b_0 Z_{t+1}$与$\{Y_n\}_{n=-\infty}^{t}$正交。继而可得最优单步预测为

$$\hat{Y}_{t+1} = -\sum_{k=1}^{m} a_k Y_{t+1-k} \tag{5.4.91}$$

易知最小均方预测误差为

$$\text{MMSE} = E\left\{\left(Y_{t+1} + \sum_{k=1}^{m} a_k Y_{t+1-k}\right)^2\right\} = E\{b_0^2 Z_{t+1}^2\} = b_0^2 E\{Z_{t+1}^2\} = b_0^2 \quad (5.4.92)$$

对于 AR(1) 序列，$b_0^2 = P(1-r^2)$，且这一结果与式(5.4.88)一致。综上所述，由 Kolmogorov-Szegö-Krein 公式(5.4.58)可得

$$b_0^2 = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\phi_Y(\omega)\mathrm{d}\omega\right\} \quad (5.4.93)$$

◀

例 5.4.4 白噪声背景下对广义马尔可夫序列的滤波、预测和平滑

考虑如下观测模型

$$Y_n = S_n + N_n, \quad n \in \mathbf{Z} \quad (5.4.94)$$

其中，$\{S_n\}_{n=-\infty}^{\infty}$和$\{N_n\}_{n=-\infty}^{\infty}$均为零均值广义平稳序列，且相互正交。假设$\{N_n\}_{n=-\infty}^{\infty}$为白序列，且 $E\{N_n^2\}=v_N^2$，$\{S_n\}_{n=-\infty}^{\infty}$为广义马尔可夫序列，且 $C_S(n)=Pr^{|n|}$，$n\in\mathbf{Z}$。参考例 5.4.2并利用$\{S_n\}_{n=-\infty}^{\infty}$和$\{N_n\}_{n=-\infty}^{\infty}$之间的正交性，可得观测样本的频谱为

$$\begin{aligned}\phi_Y(\omega) = \phi_S(\omega) + \phi_N(\omega) &= \frac{P(1-r^2)}{(1-re^{-i\omega})(1-re^{i\omega})} + v_N^2 \\ &= \frac{P(1-r^2) + v_N^2(1-re^{-i\omega})(1-re^{i\omega})}{(1-re^{-i\omega})(1-re^{i\omega})}\end{aligned} \quad (5.4.95)$$

易知，上述频谱为有理谱。

ϕ_Y 中分母多项式已经按 $A(z)A(1/z)$ 的形式进行了因式分解，其中 $A(z)=1-rz^{-1}$。分子多项式为 $N(z)=n_1z+n_0+n_1z^{-1}$，其中 $n_0=P(1-r^2)+v_N^2(1+r^2)$且 $n_1=-v_N^2r$。利用二次公式有

$$N(z) = n_1 z^{-1}(z-z_1)(z-1/z_1)$$

且

$$z_1 = -\left[(n_0^2-4n_1^2)^{1/2}+n_0\right]/2n_1$$

由于$|z_1|>1$，因此 $N(z)=B(z)B(1/z)$，其中

$$B(z) = \sqrt{-n_1/z_1}(z^{-1}-z_1) = b_0 + b_1 z^{-1} \quad (5.4.96)$$

且 $b_0=-z_1\sqrt{-n_1/z_1}$，$b_1=\sqrt{-n_1/z_1}$。相应的白化滤波器传递函数为

$$\frac{1}{\phi_Y^+(\omega)} = \frac{A(e^{i\omega})}{B(e^{i\omega})} = \frac{1-re^{-i\omega}}{b_0+b_1e^{-i\omega}} \quad (5.4.97)$$

同例 5.4.1，期望对信号序列$\{S_n\}_{n=-\infty}^{\infty}$在$(t+\lambda)$时刻的值进行估计，即 $X_t=S_{t+\lambda}$。相应的互功率谱为[参考式(5.4.24)]

$$\phi_{XY}(\omega) = e^{i\omega\lambda}\phi_S(\omega) = \frac{P(1-r^2)e^{i\omega\lambda}}{A(e^{i\omega})A(e^{-i\omega})} \quad (5.4.98)$$

将式(5.4.97)和式(5.4.98)代入式(5.4.61)，可得最优滤波器的传递函数为

$$\begin{aligned}H(\omega) &= \frac{A(e^{i\omega})}{B(e^{i\omega})}\left[\frac{P(1-r^2)e^{i\omega\lambda}}{A(e^{i\omega})B(e^{-i\omega})}\right]_+ \\ &= \left\{\frac{1-re^{-i\omega}}{b_0+b_1e^{-i\omega}}\right\}\left[\frac{P(1-r^2)e^{+i\omega\lambda}}{(1-re^{-i\omega})(b_0+b_1e^{i\omega})}\right]_+\end{aligned} \quad (5.4.99)$$

为简化式(5.4.99)，引入下列关于复变量 z 的函数

$$\hat{H}(z) = \left[\frac{P(1-r^2)}{1-rz^{-1}}\right]\left(\frac{1}{b_0+b_1z}\right) \quad (5.4.100)$$

继而利用部分分式展开可得

$$\hat{H}(z) = \frac{k'}{1-rz^{-1}} + \frac{k'}{1-z/z_1} \tag{5.4.101}$$

其中，$z_1 = -b_0/b_1$ 且 $k' = P(1-r^2)/(b_0+b_1 r)$。根据等比级数可将 $\hat{H}$ 展开为

$$\hat{H}(z) = k'\sum_{n=0}^{\infty} r^n z^{-1} + k'\sum_{n=0}^{\infty} z_1^{-1} z^n = k'\sum_{n=0}^{\infty} r^n z^{-1} + k'\sum_{n=-\infty}^{0} z_1^n z^{-1} \tag{5.4.102}$$

则 $\hat{H}$ 对应的冲激响应为

$$\hat{h}_n = \begin{cases} k' z_1^n, & n < 0 \\ 2k', & n = 0 \\ k' r^n, & n > 0 \end{cases} \tag{5.4.103}$$

故 $e^{i\omega\lambda}\hat{H}$ 对应的冲激响应为

$$\hat{h}_{n+\lambda} = \begin{cases} k' z_1^{\lambda} z_1^n, & n < -\lambda \\ 2k', & n = -\lambda \\ k' r^{\lambda} r^n, & n > -\lambda \end{cases} \tag{5.4.104}$$

滤波器 $\hat{h}_{n+\lambda}$ 如图 5.4.3 所示。为得到 $\widetilde{H}(\omega) \triangleq [e^{i\omega\lambda}\hat{H}(\omega)]_+$，可将 $\hat{h}_{n+\lambda}$ 截断为因果序列。由式(5.4.104)可得，当 $\lambda > 0$ 时，截断后的冲激响应为

图 5.4.3　滤波器 $e^{i\omega\lambda}\hat{H}(\omega)$ 的冲激响应

$$\widetilde{h}_n = \begin{cases} 0, & n < 0 \\ k' r^{\lambda} r^n, & n \leqslant 0 \end{cases} \tag{5.4.105}$$

当 $\lambda \leqslant 0$ 时

$$\widetilde{h}_n = \begin{cases} 0, & n < 0 \\ k' z_1^{\lambda} z_1^n, & 0 \leqslant n < -\lambda \\ 2k', & n = -\lambda \\ k' r^{\lambda} r^n, & n > -\lambda \end{cases} \tag{5.4.106}$$

$\lambda < 0$ 及 $\lambda > 0$ 时的滤波器分别如图 5.4.4 和图 5.4.5 所示。

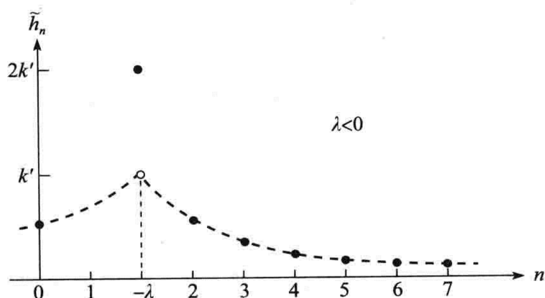


图 5.4.4　当 $\lambda < 0$ 时滤波器 $[e^{i\omega\lambda}\hat{H}(\omega)]_+$ 的冲激响应(平滑)

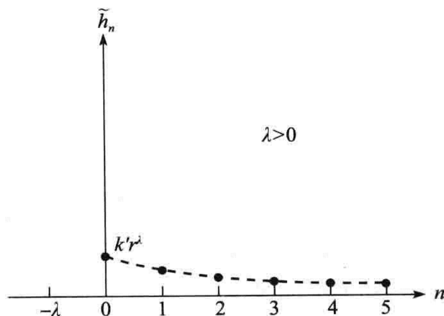


图 5.4.5　当 $\lambda > 0$ 时滤波器 $[e^{i\omega\lambda}\hat{H}(\omega)]_+$ 的冲激响应(预测)

进一步地，对预测滤波器($\lambda > 0$)进行分析。由式(5.4.105)可得

$$\widetilde{H}(e^{i\omega}) = \sum_{n=0}^{\infty} \widetilde{h}_n e^{-in\omega} = \sum_{n=0}^{\infty} k' r^{\lambda} r^n e^{-in\omega} = \frac{k' r^{\lambda}}{1 - re^{-in\omega}} \tag{5.4.107}$$

联立式(5.4.99)和式(5.4.107)，易知最优预测滤波器的传递函数为

$$H(\omega) = \frac{k' r^{\lambda}}{b_0 + b_1 e^{-i\omega}} = \frac{k' r^{\lambda}/b_0}{1 - e^{-i\omega}/z_1} \tag{5.4.108}$$

因此，最优预测器对应的冲激响应为

$$h_n = \begin{cases} 0, & n < 0 \\ \dfrac{k' r^{\lambda}}{b_0} z_1^{-n}, & n \geqslant 0 \end{cases} \tag{5.4.109}$$

等价地，也可采用下列递归公式对最优预测器进行描述

$$\hat{S}_{t+\lambda|t} = \frac{1}{z_1} \hat{S}_{t-1+\lambda|t-1} + \frac{k' r^{\lambda}}{b_0} Y_t, \quad t \in \mathbf{Z} \tag{5.4.110}$$

其中，$\hat{S}_{t+\lambda|t}$ 和 $\hat{S}_{t-1+\lambda|t-1}$ 分别表示基于 $\{Y_n\}_{n=-\infty}^{t}$ 对 $S_{t+\lambda}$ 的最优预测和基于 $\{Y_n\}_{n=-\infty}^{t-1}$ 对 $S_{t-1+\lambda}$ 的最优预测。当 $v_N^2=0$ 时(无噪条件下)，由 $\hat{S}_{t+\lambda|t}$ 可得例 5.4.2 中的纯预测结果。

下面进一步讨论单步预测滤波器($\lambda=1$)。由代数变换易得 $z_1^{-1}=r-k'r/b_0$，代入式(5.4.110)可得

$$\hat{S}_{t+1|t} = r\hat{S}_{t|t-1} + \frac{k' r}{b_0}(Y_t - \hat{S}_{t|t-1}) \tag{5.4.111}$$

上式与 5.2 节所述 Kalman-Bucy 预测滤波器类似，即通过相同方式在状态空间模型中对单步预测结果进行更新。事实上，此时的 $\{S_n\}_{n=-\infty}^{\infty}$ 为一个 AR(1)序列，因此可由下列递推公式得到[参考式(5.4.89)]

$$S_{n+1} = rS_n + [P(1-r^2)]^{1/2} W_n, \quad n \in \mathbf{Z} \tag{5.4.112}$$

其中，$\{W_n\}_{n=-\infty}^{\infty}$ 为单位方差的白序列($\{W_n\}_{n=-\infty}^{\infty}$ 为信号的预白化结果)。观测模型为

$$Y_n = S_n + N_n, \quad n \in \mathbf{Z} \tag{5.4.113}$$

其中，$\{N_n\}_{n=-\infty}^{\infty}$ 是一个方差为 v_N^2 的白序列。由于 $\{N_n\}_{n=-\infty}^{\infty}$ 与 $\{S_n\}_{n=-\infty}^{\infty}$ 和 $\{W_n\}_{n=-\infty}^{\infty}$ 均正交，因此式(5.4.112)与式(5.4.113)均为正交白噪声条件下的标量时不变 Kalman-Bucy 模型。所以，由例 5.3.2 可知，基于 Kalman-Bucy 滤波器可得已知 $\{Y_n;\ n \leqslant t\}$ 条件下对 S_t 和 S_{t+1} 的线性最小均方误差估计。与 5.2 节中的情况不同，式(5.4.112)和式(5.4.113)均为平稳模型或稳态模型。因此，相应的 Kalman-Bucy 预测滤波器为例 5.2.2 中推导结果对应的稳态形式，且与式(5.4.111)中对等效参数的有效识别相一致。此外，例 5.2.2 中的 Kalman-Bucy 模型参数(a，b，c，q，r)在这里被设定为(r，$\sqrt{P(1-r^2)}$，1，1，v_N^2)。

因此，在标量时不变情况下，可以把 Wiener-Kolmogorov 滤波器看作 Kalman-Bucy 滤波器的稳态形式，或者相应地，将 Kalman-Bucy 滤波器看作包含瞬态的 Wiener-Kolmogorov 滤波器。事实上，对于白噪声背景下的有理谱信号而言，平稳时不变的 Kalman-Bucy 滤波器与 Wiener-Kolmogorov 滤波器之间的上述关系总可以得到类似证明[详见 Anderson 和 Moore(1979)给出的进一步讨论]。此外，上述两种滤波器也可适用于其他模型，例如，Wiener-Kolmogorov 滤波可适用于更普遍意义下的信噪频谱模型，而 Kalman-Bucy 滤波也可适用于时变和非平稳状态空间模型。 ◀

5.5 习题

1. 直接证明(即不利用结论 $\hat{\boldsymbol{X}}_{t|t-1}=E\{\boldsymbol{X}_t \mid \boldsymbol{Y}_0^{t-1}\}$ 和 $\hat{\boldsymbol{X}}_{t|t}=E\{\boldsymbol{X}_t \mid \boldsymbol{Y}_0^{t}\}$)Kalman 滤波器产生的滤波误差和预测误差均与数据正交，即证明

$$E\{(\boldsymbol{X}_t-\hat{\boldsymbol{X}}_{t|t})\boldsymbol{Y}_k^T\}=\boldsymbol{0},\quad 0\leqslant k\leqslant t \tag{5.5.1}$$

和

$$E\{(\boldsymbol{X}_t-\hat{\boldsymbol{X}}_{t|t-1})\boldsymbol{Y}_k^T\}=\boldsymbol{0},\quad 0\leqslant k\leqslant t-1 \tag{5.5.2}$$

其中，$\boldsymbol{0}$ 表示全零矩阵。

2. 假设 Kalman-Bucy 模型中的状态方程被改写为如下形式

$$\boldsymbol{X}_{k+1}=F_k\boldsymbol{X}_k+G_k\boldsymbol{U}_k+\Gamma_k\boldsymbol{s}_k$$

其中，$\{\boldsymbol{s}_k\}_{k=0}^{\infty}$ 为一组已知向量，$\{\Gamma_k\}_{k=0}^{\infty}$ 为一组具有适当维度的已知矩阵(例如，$\{\boldsymbol{s}_k\}_{k=0}^{\infty}$ 可以是一组控制量)。计算相应的 Kalman-Bucy 递归表达式。

3. 当每个 $\boldsymbol{s}_k$ 均可表示为过去测量值的函数时，重做习题 2，即 $\boldsymbol{s}_k$ 可表示为 $\boldsymbol{Y}_0^k$ 的函数(例如，$\{\boldsymbol{s}_k\}_{k=0}^{\infty}$ 可以是一组反馈控制量)。

4. 假设回到对原始 Kalman-Bucy 模型的讨论，但允许利用状态变量和测量噪声间的相关性，即除下列条件外其他条件均如前所述

$$\mathrm{cov}(\boldsymbol{U}_k,\boldsymbol{V}_l)=\begin{cases}\boldsymbol{C}_k, & k=l\\ \boldsymbol{0}, & k\neq l\end{cases}$$

其中，$\boldsymbol{C}_k$ 为具有合适维度的矩阵。证明：Kalman 预测可由下列方程给出

$$\hat{\boldsymbol{X}}_{t+1|t}=\boldsymbol{F}_t\hat{\boldsymbol{X}}_{t|t-1}+\boldsymbol{K}_t(\boldsymbol{Y}_t-\boldsymbol{H}_t\hat{\boldsymbol{X}}_{t|t-1})$$

且

$$\hat{\boldsymbol{X}}_{0|-1}=\boldsymbol{m}_0$$

其中

$$\boldsymbol{K}_t=(\boldsymbol{F}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}_t^T+\boldsymbol{G}_t\boldsymbol{C}_t)(\boldsymbol{H}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}_t^T+\mathbf{R}_t)^{-1}$$

和

$$\boldsymbol{\Sigma}_{t+1|t}=\boldsymbol{F}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{F}_t^T-\boldsymbol{K}_t(\boldsymbol{F}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{H}_t^T+\boldsymbol{G}_t\boldsymbol{C}_t)+\boldsymbol{G}_t\boldsymbol{Q}_t\boldsymbol{G}_t^T$$

且

$$\boldsymbol{\Sigma}_{0|-1}=\mathrm{cov}(\boldsymbol{X}_0)$$

5. 考虑如下模型(X_k 为标量)

$$X_{k+1}=\frac{1}{2}X_k+U_k,\quad k=0,1,\cdots$$

$$Y_k=\Theta X_k+V_k,\quad k=0,1,\cdots$$

其中，U_0，U_1，…为独立同分布的 $\mathcal{N}(0,q)$ 随机变量，V_0，V_1，…为独立同分布的 $\mathcal{N}(0,r)$ 随机变量，X_0 为一个 $\mathcal{N}(0,\Sigma_0)$ 随机变量，且所有 U_k、V_k 和 X_0 均相互独立。

(a) 假设 $\Theta\equiv1$ 且 $r=q=1$。求时不变最优预测滤波器($\hat{X}_{t+1|t}$)对应的初始状态方差 Σ_0。写出此种情况下 $\hat{X}_{t+1|t}$ 的递归表达式。此时的均方预测误差是多少?

(b) 假设 $r=q=1$ 且 Σ_0 由(a)部分得到。根据 $k=0,1,\cdots,n$ 时的观测样本，确定 $\Theta=0$ 或 $\Theta=1$ 的最优判决器形式。

6. 考虑具有状态 $\boldsymbol{X}_k$ 和观测样本 $\boldsymbol{Y}_k$ 的标准 Kalman-Bucy 模型。假设 $0\leqslant j\leqslant t$ 并基于 $\boldsymbol{Y}_0^t$ 估计 $\boldsymbol{X}_j$。该估计问题的递归表达式如下所示(基于变量 t)

$$\hat{\boldsymbol{X}}_{j|t}=\hat{\boldsymbol{X}}_{j|t-1}+\boldsymbol{K}_t^a(\boldsymbol{Y}_t-\boldsymbol{H}_t\hat{\boldsymbol{X}}_{t|t-1})$$

其中

$$\boldsymbol{K}_t^a=\boldsymbol{\Sigma}_{t|t-1}^a\boldsymbol{H}_t^T[\boldsymbol{H}_t\boldsymbol{\Sigma}_{t|t-1}^a\boldsymbol{H}_t^T+\mathbf{R}_t]^{-1}$$

和

$$\boldsymbol{\Sigma}_{t+1|t}^a=\boldsymbol{\Sigma}_{t|t-1}^a[\boldsymbol{F}_t-\boldsymbol{K}_t\boldsymbol{H}_t]^T$$

且

$$\boldsymbol{\Sigma}_{j|j-1}^a=\boldsymbol{\Sigma}_{j|j-1}$$

$\boldsymbol{H}_t$、$\hat{\boldsymbol{X}}_{t|t-1}$、$\boldsymbol{\Sigma}_{t|t-1}$、$\mathbf{R}_t$、$\boldsymbol{F}_t$ 和 $\boldsymbol{K}_t$ 均与单步预测问题时相同。

(a) 证明 $\boldsymbol{\Sigma}_{t|t-1}^a=E\{(\boldsymbol{X}_j-\hat{\boldsymbol{X}}_{j|t-1})\boldsymbol{X}_t^T\}$。

(b) 证明 $\hat{\boldsymbol{X}}_{j|t}=E\{\boldsymbol{X}_j \mid \boldsymbol{Y}_0^t\}$。

7. 假设 $\boldsymbol{\Theta}$，$\boldsymbol{Z}_1$，$\boldsymbol{Z}_2$，…，$\boldsymbol{Z}_n$ 为联合高斯随机向量，$\boldsymbol{Z}_1$，$\boldsymbol{Z}_2$，…，$\boldsymbol{Z}_n$ 相互独立且均值为零。证明：已知 $\boldsymbol{Z}_1$，$\boldsymbol{Z}_2$，…，$\boldsymbol{Z}_n$ 时对 $\boldsymbol{\Theta}$ 的最小均方估计误差由下式给出

$$\hat{\boldsymbol{\theta}}=\sum_{k=1}^{n}E\{\boldsymbol{\Theta}\mid\boldsymbol{Z}_k\}-(n-1)E\{\boldsymbol{\Theta}\}$$

8. 利用习题 7 的结论推导 Kalman-Bucy 测量更新公式(5.2.14a)。

9. 考虑如下观测模型：

$$Y_k=N_k+\Theta s_k,\quad k=1,2,\cdots$$

其中，N_1，N_2，…为独立同分布 $\mathcal{N}(0,\sigma^2)$ 随机变量，$\Theta\sim\mathcal{N}(\mu,\nu^2)$ 与 N_1，N_2，…相互独立且

s_1，s_2，…为已知序列。令$\hat{\theta}_n$ 表示已知 Y_1，…，Y_n 时对 Θ 的最小均方估计误差。将该问题作为 Kalman 滤波问题，给出$\hat{\theta}_n$ 和最小均方估计误差 $E\{(\hat{\theta}_n-\Theta)^2\}$的递归表达式。

10. 考虑任意随机变量$\{X_k\}_{k=0}^{\infty}$，取值均为 0 或 1。假设对于所有的整数 $k\geqslant 1$ 和随机序列$\{x_k\}_{k=0}^{\infty}$，上述序列的概率分布为

$$P(X_k=x_k \mid X_0=x_0,\cdots,X_{k-1}=x_{k-1})$$
$$=P(X_k=x_k \mid X_{k-1}=x_{k-1}) \triangleq p_{x_k,x_{k-1}}$$

考虑观测模型

$$Y_k=X_k+N_k,\quad k=0,1,2,\cdots$$

其中，$\{N_k\}_{k=0}^{\infty}$为一组独立同分布的随机变量，与$\{X_k\}_{k=0}^{\infty}$独立且具有边缘概率密度函数 f。对每个整数 $t\geqslant 0$，令$\hat{X}_{t\mid t}$表示已知观测样本 Y_0，Y_1，…，Y_t 时对 X_t 的最小均方估计误差，$\hat{X}_{t\mid t-1}$表示已知观测样本 Y_0，Y_1，…，Y_{t-1}时对 X_t 的最小均方估计误差，且$\hat{X}_{0\mid -1}=E\{X_0\}$。证明：$\hat{X}_{t\mid t}$和$\hat{X}_{t\mid t-1}$满足下列联合递归公式

$$\hat{X}_{t|t}=\frac{\hat{X}_{t|t-1}f(y_t-1)}{\hat{X}_{t|t-1}f(y_t-1)+(1-\hat{X}_{t|t-1})f(y_t)},\quad t\geqslant 0$$

和

$$\hat{X}_{t+1|t}=p_{1,1}\hat{X}_{t|t}+p_{1,0}(1-\hat{X}_{t|t}),\quad t\geqslant 0$$

且初始条件为$\hat{X}_{0\mid -1}=P(X_0=1)$。能否将该结论推广到满足上述性质的非二进制序列？

11. 证明：式(5.3.26)～式(5.3.29)给出的 Levinson 算法可用于求解 Yule-Walker 方程(5.3.25)。

12. 考虑例 5.4.1 中的模型在 $\lambda=0$ 时的情况。对 S_t 的估计如下

$$\hat{S}_t=\sum_{n=-\infty}^{\infty}h_{t-n}Y_n$$

证明：

$$E\{(\hat{S}_{t+\lambda}-S_{t+\lambda})^2\}=\frac{1}{2\pi}\int_{-\pi}^{\pi}|1-H(\omega)|^2\phi_S(\omega)\mathrm{d}\omega+\frac{1}{2\pi}\int_{-\pi}^{\pi}|H(\omega)|^2\phi_N(\omega)\mathrm{d}\omega$$

其中，H 为$\{h_n\}_{k=-\infty}^{\infty}$对应的传递函数，并对上述误差公式中右边两项予以说明。

13. 传递函数的滤波器如下

$$H(\omega)=\frac{\phi_S(\omega)}{\phi_S(\omega)+\phi_N(\omega)}$$

证明该式为习题 12 中误差表达式对应的最小值。

第6章 连续时间信号检测

6.1 引言

在前面的章节中，假设观测集为向量集或离散集，并详细论述了信号检测和估计的基本理论。不过，无论是求解假设检验中的似然比，还是计算贝叶斯估计中的后验参数，或者寻求非贝叶斯参数估计中的最大无偏估计和最小方差无偏估计，密度族函数$\{p_\theta;\ \theta\in\Lambda\}$在整个分析过程中均扮演了重要角色。然而，正是由于在观测空间上需要指定密度族函数，不可避免地限制了观测集的形式。本章采用概率密度函数和概率分布函数并对其进行积分来解决上述问题⊖。

虽然离散时间随机过程在实际应用中占有很大的比重，但是同样可以发现有相当多的应用适用于连续时间随机过程模型。也就是说，获得的整个观测Y是随机变量$\{Y_t;\ t\in[0,\ T]\}$的集合，其中t是连续变化的参数，观测间隔为$[0,\ T]$($T>0$)。在本章和下一章中，将考虑这种观测形式下的信号检测和估计问题。本章重点考虑信号检测，而估计问题留在下一章解决。

在连续时间问题中，观测集Γ是一个全部元素都为连续时间波形的集合，这样的集合称为函数空间。为了在这样的集合中建立信号检测和估计模型，必须在这种集合下建立密度族。既然对密度函数通过积分求和后可得到(信号)概率，那么对于连续时间信号密度函数同样需要建立一种在函数空间中进行积分的方法。对这类积分方法的分析超出了常规微积分方法理论的范畴，为此，在讨论连续时间中的信号检测和估计的具体方法之前，首先给出针对此类问题的分析方法。

在6.2节中，首先简要论述抽象空间上的积分理论，其目的不在于详细说明该理论的细节，而是指出如何将密度函数的概念推广到函数空间，继而讨论在何种模型下能够有效处理连续时间信号。6.2节还将给出连续时间随机过程的Karhunen-Loéve展开模型，该模型使得整个处理方法简化为等价的离散时间过程，是本章后续部分中分析信号检测问题的关键。6.3节和6.4节将讨论具体的连续观测信号检测问题。一旦指定合适的密度函数族后，此类问题和前面所讨论的离散问题在理论上可以等价。这部分内容将主要考虑用特定的方法为实际应用中需要考虑的模型推导出适当的密度族函数，同时给出检测器结构并分析检测性能。在6.3节中将讨论高斯噪声下确知(相干)信号的检测问题，在6.4节中将讨论高斯噪声中的随机信号检测问题。

⊖ 5.4节中的线性估计问题是一个特例，因为对于这个问题仅需要知道二阶统计量。事实上，可以将观测集拓展到无限序列集。

6.2 数学基础

6.2.1 函数空间中的密度函数

为了考虑连续时间观测下的信号检测和估计问题，不能仅将观测集 Γ 限制为欧几里得几何集或是离散集，而需要将 Γ 推广到任意集合，相应的事件类 $\mathcal{G}$ 是 Γ 的任意 σ 代数子集。在这一节中，将给出任意观测集下的密度函数概念，并讨论这些概念与前面章节中针对特定观测空间所定义的相关概念的联系。本章的目的是使读者从本质上理解此类密度函数的意义，而不是给出理论的完整推导过程。

为了在函数空间中详细描述密度函数，需要给出该空间下的积分形式。从定义任意空间中的积分出发，首先介绍测度的相关概念。

定义 6.2.1 测度

如果满足下面的两条性质，则函数 $\mu:\mathcal{G}\to[0,\infty]$ 是在 $(\Gamma,\mathcal{G})$ 上的一个测度。

1) $\mu(\phi)=0$，式中 ϕ 表示零集；

2) 如果 G_1，G_2，…是在 $\mathcal{G}$ 中的一系列不相交集($G_iG_j=\phi$，$i\neq j$)，那么 $\mu(\bigcup\limits_{i=1}^{\infty}G_i)=\sum\limits_{i=1}^{\infty}\mu(G_i)$ 。

概率测度(或概率分布)是测度最常见的例子，它还满足附加条件 $\mu(\Gamma)=1$[该条件与性质 1)和性质 2)一起称为概率公理]。对概率测度进行积分是常规运算。特别地，假设 μ 是 $(\Gamma,\mathcal{G})$ 上的一个概率测度，X 是从 $(\Gamma,\mathcal{G})$ 到 $(\mathbf{R},\mathcal{B})$ 的可测函数[⊖]，那么在 $(\mathbf{R},\mathcal{B})$ 上与 X 对应的概率分布 P_X 可由下式决定

$$P_X(A)=\mu(X^{-1}(A)),\quad A\in\mathcal{B} \tag{6.2.1}$$

式中 $X^{-1}(A)\triangleq\{y\in\Gamma\,|\,X(y)\in A\}$。

函数 X 的值对应随机变量，期望 $E\{X\}$ 可以采用传统方式定义，其大小可认为是实变量 x 经过分布 P_X 加权后的平均值；同时，也可以认为它是 $X(y)$ 经过在 $(\Gamma,\mathcal{G})$ 上的概率测度 μ 加权后的平均值，即

$$E\{X\}=\int_{\Gamma}X(y)\mu(\mathrm{d}y)\equiv\int X\mathrm{d}\mu \tag{6.2.2}$$

显然，$E\{X\}$ 可理解为 X 关于 μ 的积分。假设 X_1 和 X_2 分别是 $(\Gamma,\mathcal{G})$ 上的两个函数，并已知均值 $E\{X_1\}$ 和 $E\{X_2\}$，对于任意标量 α 和 β，利用期望性质可以证明 $E\{\alpha X_1+\beta X_2\}=\alpha E\{X_1\}+\beta E\{X_2\}$ 成立。采用上面的记法，则有

$$\int(\alpha X_1+\beta X_2)\mathrm{d}\mu=\alpha\int X_1\mathrm{d}\mu+\beta\int X_2\mathrm{d}\mu \tag{6.2.3}$$

即式(6.2.2)是 X 的线性函数，这也是积分的基本要求。

⊖ 一个函数 X 在 $(\Gamma,\mathcal{G})$ 到 $(\mathbf{R},\mathcal{B})$ 上可测是指：对于任何一个满足 $A\in\mathcal{B}$ 的集合，都有 $\{y\in\Gamma\,|\,X(y)\in A\}$ 在 $\mathcal{G}$ 上。这个条件意味着任何一个关于 X 在实数线上的事件，即 $X(y)$ 在 A 上，都可以映射回原始观测空间。这个条件使得可以通过 X 在 $(\mathbf{R},\mathcal{B})$ 上定义一个概率分布，如前文所述。

对于任意事件 $F\in\mathcal{G}$，可以定义一个从 Γ 到 $\mathbf{R}$ 的特殊可测函数，即

$$I_F(y)=\begin{cases}1, & y\in F\\ 0, & y\in F^c\end{cases} \tag{6.2.4}$$

正如在第 3 章中提到的，此函数称为 F 的指示函数。既然 I_F 定义了一个离散随机变量，它取值为 1 的概率为 $\mu(F)$，取值为 0 的概率为 $\mu(F^c)$，那么

$$E\{I_F\}=1\mu(F)+0\mu(F^c)=\mu(F) \tag{6.2.5}$$

或等价于

$$\int I_F\mathrm{d}\mu=\mu(F) \tag{6.2.6}$$

因此，通过适当选取 X 后，求解 $\int X\mathrm{d}\mu$ 就可得到任意集合 F 的测度 μ。由于 $\int X\mathrm{d}\mu$ 定义在 μ 上，因此积分 $\int X\mathrm{d}\mu$ 和测度 μ 是同一概念。$\int I_F\mathrm{d}\mu$ 有时也记作 $\int_F\mathrm{d}\mu$ 或者 $\int_F\mu(\mathrm{d}y)$，对于函数 X 和集合 $F\in\mathcal{G}$，记作

$$\int_\Gamma I_F(y)X(y)\mu(\mathrm{d}y)=\int_F X(y)\mu(\mathrm{d}y) \tag{6.2.7}$$

从上面的定义中可以看到，函数 X 关于概率测度 μ 的积分不过是常见的 X 的期望⊖。测度积分的概念可以很容易地从概率测度推广到其他类型的测度上。特别地，如果 $\mu(\Gamma)<\infty$，则 μ 是有限的；如果 Γ 可以写作互斥事件 Γ_1，Γ_2，…的并集，且每个事件 Γ_i 满足 $\mu(\Gamma_i)<\infty$，则 μ 是 σ 有限的。

除了归一化 $\mu(\Gamma)=1$ 外，有限测度在本质上和概率测度相同。特别地，对于任意有限测度 μ，定义概率测度 μ' 为

$$\mu'(F)=\mu(F)/\mu(\Gamma),\quad F\in\mathcal{G} \tag{6.2.8}$$

则函数 X 关于 μ 的积分可以直接定义为

$$\int X\mathrm{d}\mu=\mu(\Gamma)\int X\mathrm{d}\mu' \tag{6.2.9}$$

同理，如果 μ 是 σ 有限测度，可以将它分解为有限测度的和的形式，即

$$\mu(F)=\sum_{i=1}^{\infty}\mu_i(F),\quad F\in\mathcal{G} \tag{6.2.10}$$

式中，对于每一个 i，有

$$\mu_i(F)=\mu(F\cap\Gamma_i),\quad F\in\mathcal{G} \tag{6.2.11}$$

利用测度定义中的性质 2，有 $\mu(F\cap\Gamma_i)\leqslant\mu(\Gamma_i)\leqslant\infty$ 成立，进而对于任意 σ 有限测度，可以定义

$$\int X\mathrm{d}\mu=\sum_{i=1}^{\infty}\int X\mathrm{d}\mu_i \tag{6.2.12}$$

⊖ 值得注意的是，可以首先定义 $\int X\mathrm{d}\mu$，然后再定义 $E\{X\}=\int X\mathrm{d}\mu$[参看 Billingsley(1979)]。它比前面的方法更加合理，不过，这样做的目的是用已知的 $E\{X\}$ 来定义未知的 $\int X\mathrm{d}\mu$。

式中$\int X\mathrm{d}\mu_i$根据式(6.2.9)定义[⊖]。

在定义上述积分之后(此类积分也称为Lebesgue-Stieltjes积分)，就可以进一步给出概率密度的概念。为此，定义测度的绝对连续性如下。

定义 6.2.2 测度的绝对连续性

假设μ_0和μ_1是$(\Gamma, \mathcal{G})$上的两个测度，如果由条件$\mu_0(F)=0$可以得出$\mu_1(F)=0$，则称μ_1关于μ_0绝对连续，记为$\mu_1 \ll \mu_0$。

根据上述定义，有下面的结论成立。

命题 6.2.1 Radon-Nikodym定理

假设μ_0和μ_1是$(\Gamma, \mathcal{G})$上的两个σ有限测度，且$\mu_1 \ll \mu_0$，那么存在可测函数$f:\Gamma \to \mathbf{R}$使得

$$\mu_1(F) = \int_F f\mathrm{d}\mu_0, \quad F \in \mathcal{G} \tag{6.2.13}$$

此外，在除去$\mu_0(G_0)=0$的集合G_0上，f唯一确定。

鉴于该结论的详细证明过程在大多数关于积分和测度的书上都能找到[如Billingsley(1979)]，所以此处从略。需要注意的是：Radon-Nikodym定理说明了任意事件的μ_1测度可以通过求函数f对μ_0在F上的积分得到。函数f通常称为测度μ_1对于测度μ_0的Radon-Nikodym导数，记为$f=\mathrm{d}\mu_1/\mathrm{d}\mu_0$。下面的两个例子将进一步说明Radon-Nikodym导数在向量集和离散集上具有的形式。

例 6.2.1 向量集上的Radon-Nikodym导数

假设$(\Gamma, \mathcal{G})=(\mathbf{R}, \mathcal{B})$，存在一个定义在$(\mathbf{R}, \mathcal{B})$上的唯一测度$\mu$，满足任意间隔的$\mu$测度就是它的长度[例如，$\mu([a, b])=|b-a|$]。该测度就是众所周知的勒贝格(Lebesgue)测度，因为可将$\mathbf{R}$分解为可数个有限长区间的并集，所以σ有限。如果g是一个通常意义上的黎曼积分函数$\int_{-\infty}^{\infty} g(y)\mathrm{d}y$,那么就有$\int_{\mathbf{R}} g\mathrm{d}\mu = \int_{-\infty}^{\infty} g(y)\mathrm{d}y$。也就是说，关于勒贝格测度的积分本质上与普通微积分学的积分一样，尽管该函数勒贝格可积，但未必黎曼可积。

假设P是连续时间变量在$(\mathbf{R}, \mathcal{B})$上的概率测度，$p$表示相应的概率密度函数(pdf)。显然对于任意的$A \in \mathcal{B}$，有$P(A) = \int_A p(y)\mathrm{d}y$。又因为$\int_A p(y)\mathrm{d}y = \int_A p\mathrm{d}\mu \equiv \int_A p(y)\mu(\mathrm{d}y)$，所以有$p = \mathrm{d}P/\mathrm{d}\mu$成立。换言之，连续时间变量的概率密度函数是勒贝格测度意义上概率测度的Radon-Nikodym导数。

勒贝格测度是在$\mathbf{R}$上的“自然”测度，因为它与长度有关，而长度是与任意间隔相联系的最自然的数。同样也可以在$\mathbf{R}^2$和$\mathbf{R}^3$上定义勒贝格测度，此时，需要分别采用面积

⊖ 在定义式(6.2.12)时，必须考虑和的收敛性。当$\sum_{i=1}^{\infty}\int X^+ \mathrm{d}\mu_i = \infty$且$\sum_{i=1}^{\infty}\int X^- \mathrm{d}\mu_i < \infty$[定义$X^+(y)=\max\{X(y), 0\}$，$X^-(y)=\max\{-X(y), 0\}$]时，式(6.2.12)的和将会趋于$\infty$。同理，当$\sum_{i=1}^{\infty}\int X^+ \mathrm{d}\mu_i < \infty$且$\sum_{i=1}^{\infty}\int X^- \mathrm{d}\mu_i = \infty$时，$\int X\mathrm{d}\mu = -\infty$。当$\sum_{i=1}^{\infty}\int X^+ \mathrm{d}\mu_i$和$\sum_{i=1}^{\infty}\int X^- \mathrm{d}\mu_i$都无限时，$\int X\mathrm{d}\mu$没有定义。

和体积作为大小的自然测度。基于 $\mathbf{R}^n$ 中常见的"体积"概念，可以推广到 $\mathbf{R}^n(n>3)$。在这些情况中，每一种连续随机向量的概率测度都是与勒贝格测度有关的 Radon-Nikodym 导数。◀

例 6.2.2　离散集上的 Radon-Nikodym 导数

假设 Γ 是离散集合 $\{\gamma_1, \gamma_2, \cdots\}$，且 $\mathcal{G}=2^\Gamma$，在 $(\Gamma, \mathcal{G})$ 上定义测度 μ 为

$$\mu(F) = F\text{ 中元素的个数} \tag{6.2.14}$$

那么 μ 就是事件 F 中的元素个数，也称为计数测度。由于 Γ 是集合 $\{\gamma_i\}$ 的并集，且集合 $\{\gamma_i\}$ 只有一个元素，所以它是 σ 有限的。对于任意函数 $g:\Gamma\to\mathbf{R}$，可以很容易地得到

$$\int_\Gamma g(y)\mu(\mathrm{d}y) = \sum_{i=1}^{\infty} g(\gamma_i) \tag{6.2.15}$$

即在 Γ 上关于计数测度的积分就是简单的求和。

假设 P 是在 Γ 上概率分布函数为 p 的概率测度，那么对于任意的 $F\in\mathcal{G}$，有

$$P(F) = \sum_{\gamma_i\in F} p(\gamma_i) = \int_F p(y)\mu(\mathrm{d}y) \tag{6.2.16}$$

因此 $p=\mathrm{d}P/\mathrm{d}\mu$，即概率分布函数是计数测度的 Radon-Nikodym 导数。◀

例 6.2.1 和例 6.2.2 说明了在前面章节所用的密度函数是 Radon-Nikodym 导数的特殊情况。就所介绍的符号(即 $E\{g(Y)\}=\int_\Gamma g(y)p(y)\mu(\mathrm{d}y)$) 而言，引入测度 μ 后，在积分意义上具有更一般性的解释，进而可将前面章节得到的结论推广到更为普遍的观测空间。

所感兴趣的模型通常由 (Γ, G) 上的分布族(如概率测度) $\{P_\theta;\ \theta\in\Lambda\}$ 构成。如果存在一个 $(\Gamma, \mathcal{G})$ 上的 σ 有限测度 μ，使得对所有的 $\theta\in\Lambda$ 有 $P_\theta\ll\mu$，那么就可以构造一个密度族 $\{p_\theta;\ \theta\in\Lambda\}$，其中 $p_\theta=\mathrm{d}P_\theta/\mathrm{d}\mu$，$\theta\in\Lambda$。一旦有了关于 p_θ 更为简便的解释后，所有在假设检验和参数估计中得到的结论都可以推广到更一般的观测空间，包括第 2 章和第 4 章中关于假设检验和参数估计的一般性结论，以及 3.3 节中推导的切诺夫界。此外，如果用边缘密度代替 Radon-Nikodym 导数，在独立同分布的实值观测序列下得到的结论也能推广到一般观测空间中的独立同分布观测序列[⊖]，包括第 3 章中离散时间检测的结论和 4.4 节中最大似然估计的渐近分析结论。

对于连续时间中的信号检测和估计问题，在掌握了作为分析基础的 Lebesgue-Stieltjes 积分和 Radon-Nikodym 微分后，进一步考虑二进制参数集 $\Lambda=\{0, 1\}$ 上的一些特殊性质。

考虑 $(\Gamma, \mathcal{G})$ 上的两个概率测度 P_0 和 P_1，总存在 σ 有限测度 μ 使得 P_0 和 P_1 绝对连续。例如，易知有限测度 $\mu=P_0+P_1$ 支配着 P_0 和 P_1。为不失一般性，假设存在测度 μ，可以定义密度 $p_j\triangleq\mathrm{d}P_j/\mathrm{d}\mu$，$j=0, 1$。给定 P_0 和 P_1 后，贝叶斯、极小化极大误差和尼曼-皮尔逊最佳检验都是将似然比 $L=p_1/p_0$ 与门限 τ 进行比较，当 $L(y)$ 大于 τ 时，判决为 H_1；当 $L(y)$ 小于 τ 时，判决为 H_0；如果 $L(y)=\tau$，则进行随机判决。如果 $P_1\ll P_0$，那么 P_1 对于 P_0 可微。对满足 P_0 和 P_1 绝对连续的任意测度 μ 而言，可以得到

⊖ 这类问题出现在阵列处理等应用中。在阵列处理中，接收到 n 个观测值，每一个都可以表示为一个阵元传感器的输出波形。

$$\frac{\mathrm{d}P_1}{\mathrm{d}P_0}=\frac{\mathrm{d}P_1/\mathrm{d}\mu}{\mathrm{d}P_0/\mathrm{d}\mu}=\frac{p_1}{p_0} \tag{6.2.17}$$

因此，当 $P_1 \ll P_0$ 时，似然比仅仅是 P_1 关于 P_0 的 Radon-Nikodym 导数。

如果 $P_1 \ll P_0$，应用 Radon-Nikodym 定理，可得到

$$P_1(F)=\int_F L\,\mathrm{d}P_0,\quad F\in\mathcal{G} \tag{6.2.18}$$

更一般地，对任意一对概率测度 P_0 和 P_1，总存在一个可测函数 $f:\Gamma\to\mathbf{R}$ 和一个满足$P_0(H)=0$ 的集合 $H\in\mathcal{G}$，使得

$$P_1(F)=\int_F f\,\mathrm{d}P_0+P_1(F\cap H),\quad F\in\mathcal{G} \tag{6.2.19}$$

实际上，集合 H 可根据下式进行选择

$$H=\{y\in\Gamma\,|\,p_1(y)>0\quad \text{且}\quad p_0(y)=0\} \tag{6.2.20}$$

当 $y\in H^c$ 时，$f(y)$可选为似然比 $L(y)$；当 $y\in H$ 时，可以任意选取。这是因为 $P_0(H)=0$，f 在 H 上的取值不影响积分结果。当 $y\in H$ 时，最佳检验总是选择假设 H_1；当 $y\in H^c$ 时，则根据 $f(y)$和 τ 的比较结果进行选择。

在式(6.2.19)中，如果 $P_1(H)=0$，右边的第二项也为 0，$P_1\ll P_0$ 且 $f=\mathrm{d}P_1/\mathrm{d}P_0$。另一方面，如果 $P_1(H)=1$，那么对于所有的 $F\in\mathcal{G}$，$P_1(F\cap H)=P_1(F)$且右边第一项为 0。换言之，事件 H 在 P_1 下以概率 1 发生，事件 H^c 在 P_0 下以概率 1 发生。因此，判决准则"$y\in H$ 选择 H_1 和 $y\in H^c$ 选择 H_0"的误判概率为 0。此时，称 P_0 和 P_1 是奇异的(用符号 $P_0\perp P_1$ 表示)；若用 P_0 和 P_1 表示与此对应的信号检测模型，则称为奇异性检测。

本质上，测度间奇异意味着密度 p_0 和 p_1 的支撑集不相交。在实际模型中，对于 $\Gamma=\mathbf{R}^n$ 或 Γ 离散，很少出现奇异现象。然而，对于连续时间观测值，"奇异"是一个更为微妙的问题，它可能在其他一些看似合情合理的模型中出现。在某假设下，例如噪声过程具有有限带宽，则噪声波形可能是连续函数。此时，如果想检测加性噪声中是否包含信号，如图 6.2.1 所示，它的奇异性非常明显。显然，信号的存在会打破观测值在 $t=T/2$ 处的连续性，即可根据观测波形在 $t=T/2$处是否连续来判断有无信号。(注意，上述奇异性表述不适用于离散时间模型，因为在离散时间中不存在时间连续的概念。)虽然上述模型具有奇异性，但是在实际问题中奇异性要隐蔽得多。本质上，相比噪声而言，当信号快速变化时，在加性噪声中检测信号就会出现奇异性。尽管在检测实际物理信号时不存在奇异性，但正是因为奇异性对应模型的非理想情况，有必要研究检测模型可能存在的奇异性。针对奇异性检测问题的研究已经非常深入了，下面的章节将进行简单介绍。

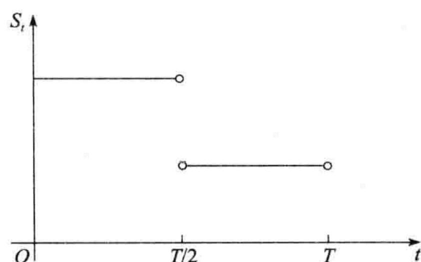


图 6.2.1　连续加性噪声中奇异性检测的信号示例

$P_1\ll P_0$和 $P_1\perp P_0$ 是式(6.2.19)的两个极端情况。当然，也可以考虑中间情况，即 $0<P_1(H)<1$。然而，对于大多数连续时间检测模型而言，情况经常是两个极端之一。事实上，有相当普遍的一类连续时间模型或是奇异，或是 P_0 和 P_1 等价，即 $P_1\ll P_0$ 和 $P_0\ll$

P_1（记为 $P_0 \equiv P_1$）。因此，连续时间检测问题的求解过程可分为以下几步：

1）确定该问题是否是奇异的；

2）如果该问题非奇异，计算似然比；

3）求解错误概率。

在介绍具体求解方法之前，为便于后续分析，先给出连续时间的表达式。

6.2.2　Grenander 定理和 Karhunen-Loéve 展开式

根据 6.2.1 节中所涉及的概念，可解释函数空间中的概率测度和似然比，但却无法求出它们在检测器或估计器中的具体表达式。因此，在本节和下节中将讨论求解这些表达式的方法。

通常可根据下述步骤计算连续时间观测模型的似然比：在连续时间观测过程 $\{Y_t;\ t\in[0,\ T]\}$ 可简化为一个相等的序列 $\{Z_1,\ Z_2,\ \cdots\}$ 的基础上，首先将无限长序列 $\{Z_k\}_{k=1}^{\infty}$ 截断成 $\{Z_1,\ \cdots,\ Z_n\}$，然后按照第 3 章中的方法计算序列 $\{Z_1,\ Z_2,\ \cdots,\ Z_n\}$ 的似然比，最后求 $n\to\infty$ 的极限，序列 $\{Z_k\}_{k=1}^{\infty}$ 的似然比将收敛到连续时间信号的似然比。实际上，正如下面将要验证的结论，在一般条件下，上述收敛性总成立。

命题 6.2.2　Grenander 定理

假设 $(\Gamma,\ \mathcal{G})=(\mathbf{R}^{\infty},\ \mathcal{B}^{\infty})$ ㊀，P_0 和 P_1 是 $(\Gamma,\ \mathcal{G})$ 上的两个概率测度。进一步假设对于每一个正整数 n，Y_1^n 是在 P_0 和 P_1 下的连续随机向量，且在 P_j 下的密度为 $p_j^n(y_1^n)$。对于每一个 n，在 Γ 上定义函数 f_n，使得

$$f_n(y)=\frac{p_1^n(y_1^n)}{p_0^n(y_1^n)},\quad y\in\Gamma \tag{6.2.21}$$

那么，在 P_0 下，$f_n(Y)\to f(Y)$ 依概率成立；在 P_1 测度下，$f_n(Y)\to f(Y)$ 在事件集 H^c 上依概率成立；

在 P_1 下，$f_n(Y)\to\infty$ 在事件集 H 上依概率成立。其中，f 和 H 分别为出现在式(6.2.19)中的函数和事件，即

$$P_1(F)=\int_F f\,\mathrm{d}P_0+P_1(F\cap H),\quad F\in\mathcal{G} \tag{6.2.22}$$

上述定理称为 Grenander 定理，具体证明过程可参见 Grenander(1981)，在此省略。Grenander 定理表明：首先截断序列，然后估计截断序列在概率意义上的似然比极限，这样就可以计算出空间中无限长序列的似然比。下一步将构造连续时间观测问题的等价观测序列。为此，首先给出一些有用的定义。

如果一个随机过程 $\{Y_t;\ t\in[0,\ T]\}$ 对于所有的 $t\in[0,\ T]$ 有 $E\{Y_t^2\}<\infty$，则称该过程为二阶过程。对于一个二阶过程，自协方差函数定义为

$$C_Y(t,u)=\operatorname{cov}(Y_t,Y_u),\quad t,u\in[0,T] \tag{6.2.23}$$

对于一个在 $[0,\ T]$ 上的实值函数 h，$\int_0^T h(t)Y_t\mathrm{d}t$ 的均方积分定义为当 $n\to\infty$ 和 $\Delta_n\triangleq$

㊀ 如 3.4 节所述，$\mathbf{R}^{\infty}$ 是全为实数的单边序列，即 $\mathbf{R}^{\infty}=\{y\,|\,y=\{y_k\}_{k=1}^{\infty}$ 且 $y_k\in\mathbf{R},\ k\geqslant1\}$。$\mathcal{B}^{\infty}$ 是 $\mathbf{R}^{\infty}$ 上 Borel 集的 σ 代数，即包含所有形如 $\{y=\{y_k\}_{k=1}^{\infty}\,|\,(y_1,\ \cdots,\ y_n)\in A\}$，$A\in\mathcal{B}^n$，$n=1,\ 2,\ \cdots$ 的集合的最小 σ 代数。

$\max\limits_{1\leqslant i\leqslant n}|t_i^{(n)}-t_{i-1}^{(n)}|\to 0$时，如下序列和的均方极限[⊖]

$$\sum_{i=0}^{n}h(\xi_i^{(n)})Y_{\xi_i^{(n)}}[t_i^{(n)}-t_{i-1}^{(n)}] \tag{6.2.24}$$

式中，对于每一个正整数 n，$0=t_0^{(n)}<t_1^{(n)}<\cdots<t_n^{(n)}=T$ 构成$[0,\ T]$上的一个子集，且对 $i=1,\ \cdots,\ n$ 有 $\xi_i^{(n)}\in[t_{i-1}^{(n)},\ t_i^{(n)}]$。注意，除了 $Y_{\xi_i^{(n)}}$ 是随机的以外，式(6.2.24)是普通形式的黎曼和，其极限为$\int_0^T h(t)Y_t\mathrm{d}t$。由于被加数是随机的，因此在随机意义上定义极限非常重要，尤其是均方极限。此外，也可用均方 Stieltjes 积分$\int_0^T h(t)\mathrm{d}Y_t$ 来定义，它是下述序列和的均方极限

$$\sum_{i=0}^{n}h(\xi_i^{(n)})(Y_{t_i^{(n)}}-Y_{t_{i-1}^{(n)}}) \tag{6.2.25}$$

式中 $t_i^{(n)}$ 和 $\xi_i^{(n)}$ 按照式(6.2.24)定义。

根据上面的定义，现在可以给出 Mercer 定理和 Karhunen-Loéve 展开式。

命题 6.2.3 Mercer 定理和 Karhunen-Loéve 展开式

假设$\{Y_t;\ t\in[0,\ T]\}$是零均值二阶随机过程，它的自协方差函数为 $C_Y(t,\ u)$，且在$[0,\ T]\times[0,\ T]\equiv[0,\ T]^2$ 上连续，那么 C_Y 可以展开为一致绝对收敛序列

$$C_Y(t,u)=\sum_{k=1}^{\infty}\lambda_k\psi_k(t)\psi_k(u),\quad (t,u)\in[0,T]^2 \tag{6.2.26}$$

式中$\{\lambda_k\}_{k=1}^{\infty}$和$\{\psi_k\}_{k=1}^{\infty}$是 C_Y 的特征值和相应的正交特征函数，即$\{\lambda_n\}_{n=1}^{\infty}$和$\{\psi_n\}_{n=1}^{\infty}$是下面的积分方程的解

$$\lambda\psi(t)=\int_0^T C_Y(t,u)\psi(u)\mathrm{d}u,\quad 0\leqslant t\leqslant T \tag{6.2.27}$$

且 $n\neq m$ 时 $\int_0^T\psi_n(t)\psi_m(t)\mathrm{d}t=0$ 且 $\int_0^T\psi_n^2(t)\mathrm{d}t=1$ 。

进而，$\{Y_t;\ t\in[0,\ T]\}$的均方收敛级数表示为

$$Y_t=\sum_{k=1}^{\infty}Z_k\psi_k(t),\quad 0\leqslant t\leqslant T \tag{6.2.28}$$

式中

$$Z_k=\int_0^T\psi_k(t)Y_t\mathrm{d}t,\quad k=1,2,\cdots \tag{6.2.29}$$

关于式(6.2.26)的有效性就是著名的 Mercer 定理，式(6.2.28)所描述的展开式就是 Karhunen-Loéve 展开式。Mercer 定理是数学分析中的经典理论，可在 Lovitt(1950)等处找到证明，Karhunen-Loéve 展开式的有效性是 Mercer 定理的推论，其证明结果可以在

⊖ 即

$$E\left\{\left[\int_o^T h(t)Y_t\mathrm{d}t-\sum_{i=1}^{n}h(\xi_i^{(n)})Y_{\xi_i}^{(n)}(t_i^{(n)}-t_{i-1}^{(n)})\right]^2\right\}\to 0$$

式(6.2.24)的极限存在当且仅当

$$\int_o^T\int_0^T h(t)h(u)R_Y(t,u)\mathrm{d}t\mathrm{d}u<\infty$$

其中 $R_Y(t,\ u)\triangleq E\{Y_tY_u\}\equiv C_Y(t,\ u)+E\{Y_t\}E\{Y_u\}$[参看 Parzen(1962)]。

Thomas(1971)中找到。注意，Mercer 定理是按照 C_Y 的特征值和正交特征函数对其进行频谱展开，正如对 $n\times n$ 的协方差矩阵 $\boldsymbol{\Sigma}$ 进行谱分解一样，即

$$\boldsymbol{\Sigma} = \sum_{k=1}^{\infty} \lambda_k \boldsymbol{v}_k \boldsymbol{v}_k^{\mathrm{T}}$$

采用 3.2 节中矩阵的特征值$\{\lambda_k\}_{k=1}^{n}$和特征向量$\{\boldsymbol{v}_k\}_{k=1}^{n}$表示。(这两个表达式都是算子理论中普通表达式的特殊例子，因此它们完全等价。)同样，Karhunen-Loéve 展开式类似于将 n 维零均值随机向量 $\boldsymbol{Y}$ 分解为

$$\boldsymbol{Y} = \sum_{k=1}^{\infty} Z_k \boldsymbol{v}_k \tag{6.2.30}$$

式中 $Z_k = \boldsymbol{v}_k^{\mathrm{T}} \boldsymbol{Y}$。在具体应用 Karhunen-Loéve 展开式时，会进一步讨论上述相似性。

Karhunen-Loéve 展开式提供了在连续时间过程中分离随机性和时变性的方法，即$\{Y_t;\ t\in[0,\ T]\}$的随机性归结到序列$\{Z_k\}_{k=1}^{\infty}$中，而该过程的时变性体现在函数序列$\{\psi_k\}_{k=1}^{\infty}$中。式(6.2.28)通过组合上述两者来表示该过程。一方面，在$\{Z_k\}_{k=1}^{\infty}$的基础上，根据式(6.2.28)可确定$\{Y_t;\ t\in[0,\ T]\}$；另一方面，根据式(6.2.29)由$\{Y_t;\ t\in[0,\ T]\}$可计算出$\{Z_k\}_{k=1}^{\infty}$。因此，$\{Z_k\}_{k=1}^{\infty}$是连续时间过程$\{Y_t;\ t\in[0,\ T]\}$的等价观测序列。换言之，Karhunen-Loéve 展开式可将连续时间过程转换为一个等价序列。

需要注意的是，假设 C_Y 在$[0,\ T]^2$ 上连续是非常合理的。容易证明，这个条件与$\{Y_t;\ t\in[0,\ T]\}$均方连续的条件等价，即

$$\lim_{u\to t} E\{(Y_u - Y_t)^2\} = 0, \quad t\in[0,T] \tag{6.2.31}$$

[参见 Parzen(1962)]。$\{Y_t;\ t\in[0,\ T]\}$是零均值过程的条件并不失一般性，这是因为即使$\{Y_t;\ t\in[0,\ T]\}$的均值不等于零，利用定理也可对零均值过程$\{Y_t - E(Y_t);\ t\in[0,\ T]\}$进行展开。由于协方差 C_Y 不会随均值的变化而变化，因此不会改变在展开式中用到的特征函数。然而，在式(6.2.29)的右边需要加入 $E\{Y_t\}$项，即式(6.2.29)需要改写为

$$Z_k = \int_0^T \psi_k(t)(Y_t - E\{Y_t\})\mathrm{d}t, \quad k = 1,2,\cdots \tag{6.2.32}$$

从表面上看，Karhunen-Loéve 展开式和 Grenander 定理可用于解决任何一个连续时间观测下的信号检测问题。然而，事实并非如此，因为该方法面临两大问题。一是计算 Karhunen-Loéve 展开式需要通过自协方差函数 C_Y 得到$\{Y_t;\ 0\leqslant t\leqslant T\}$的概率分布。二是计算式(6.2.29)中的序列$\{Z_k\}_{k=1}^{\infty}$时需要已知 C_Y 的特征函数。以二元假设检验问题为例，并不知道在两个可能的分布中哪一个是有效的先验信息。因此，除非$\{Y_t;\ t\in[0,\ T]\}$的自协方差函数在两个假设下具有相同的特征函数，否则无法求解出正确的等价观测序列$\{Z_k\}_{k=1}^{\infty}$。

尽管在 6.3 节中有很大一类问题都可以避免出现前面的情况，但在应用 Grenander 定理计算似然比时仍会面临更为棘手的问题。假设 C_Y 的特征函数在两种假设下是相同的，因此可以根据下式产生序列$\{Z_k\}_{k=1}^{\infty}$

$$Z_k = \int_0^T \psi_k(t) Y_t \mathrm{d}t, \quad k = 1,2,\cdots \tag{6.2.33}$$

在应用 Grenander 定理时，对所有的 n，都需要已知 Z_1^n 在每个假设下的概率密度。即使存

在一些特殊的例子，求解随机变量 Z_k 的边缘密度也是极其困难的。若没有进一步的假设，就无法继续求解似然比。不过，如果随机过程 $\{Y_t;\ t\in[0,\ T]\}$ 是高斯的，那么根据 Grenander 定理来求解似然比就容易得多。鉴于许多物理现象(包括电路中的热噪声)可以用高斯过程进行精确建模，所以高斯随机过程的应用十分广泛，这也是本节对其进行详细讨论的原因。

定义高斯随机过程的最简单方式是：如果随机过程 $\{Y_t;\ t\in[0,\ T]\}$ 的采样向量全是高斯随机向量，那么该过程就是高斯的。也就是说，具有 $(Y_{t_1},\ Y_{t_2},\ \cdots,\ Y_{t_n})^{\mathrm{T}}$ 形式的所有向量都应该满足多元高斯分布，其中 n 是正整数且 $t_i\in[0,\ T]$，$i=1,\ \cdots,\ n$。类似于高斯随机向量，高斯随机过程的线性变换也满足高斯性质。鉴于积分是线性运算，且在 Karhunen-Loéve 展开式中的序列是通过 $\{Y_t;\ t\in[0,\ T]\}$ 的线性变换得到的，所以对于每一个 n，Z_1^n 构成一个高斯随机向量，它的密度由随机变量 $Z_1,\ \cdots,\ Z_n$ 的均值和协方差确定。

在计算高斯过程的均值和协方差之前，首先可以交换期望运算和均方积分的顺序，因此有

$$E\{Z_k\}=E\left\{\int_0^T\psi_k(t)Y_t\mathrm{d}t\right\}=\int_0^T\psi_k(t)E\{Y_t\}\mathrm{d}t=0 \tag{6.2.34}$$

式中假设 $\{Y_t;\ t\in[0,\ T]\}$ 满足零均值条件。同样，协方差可表示为

$$\begin{aligned}\operatorname{cov}(Z_k,Z_m)&=E\{Z_kZ_m\}=E\left\{\int_0^T\psi_k(t)Y_t\mathrm{d}t\int_0^T\psi_m(u)Y_u\mathrm{d}u\right\}\\&=E\left\{\int_0^T\int_0^T\psi_k(t)\psi_m(u)Y_tY_u\mathrm{d}t\mathrm{d}u\right\}\\&=\int_0^T\int_0^T\psi_k(t)\psi_m(u)E\{Y_tY_u\}\mathrm{d}t\mathrm{d}u\\&=\int_0^T\int_0^T\psi_k(t)\psi_m(u)C_Y(t,u)\mathrm{d}t\mathrm{d}u\end{aligned} \tag{6.2.35}$$

对上式右边的 u 进行积分，结合式(6.2.27)，可以得到

$$\operatorname{cov}(Z_k,Z_m)=\lambda_m\int_0^T\psi_k(t)\psi_m(u)\mathrm{d}t \tag{6.2.36}$$

再利用 $\{\psi_k\}_{k=1}^{\infty}$ 的正交性，有

$$\operatorname{cov}(Z_k,Z_m)=\begin{cases}\lambda_k, & k=m\\0, & k\neq m\end{cases} \tag{6.2.37}$$

式(6.2.37)表明，当 $k\neq m$ 时，Z_k 和 Z_m 不相关，又由于它们都是高斯的，所以相互独立。因此，当 $\{Y_t;\ t\in[0,\ T]\}$ 是零均值高斯过程时，Karhunen-Loéve 系数 $\{Z_k\}_{k=1}^{\infty}$ 是服从 $Z_k\sim\mathcal{N}(0,\ \lambda_k)$ 分布[⊖]的独立随机变量序列。

6.3　高斯噪声中的确知信号和部分参数确知信号检测

在明确 6.2 节中的分析方法之后，本节重点考虑连续时间观测值下信号检测的具体问题。检测的基本方法是首先通过 Karhunen-Loéve 展开式将连续时间观测集简化为一个等

⊖ 值得注意的是：式(6.2.37)的协方差结构并不依赖于高斯假设，且对于 Karhunen-Loéve 展开式的系数总成立。也正是因为系数具有非相关特性，使得 Karhunen-Loéve 展开式在随机过程的众多类傅里叶展开中具有非常重要的地位。

价的离散时间集，然后应用 Grenander 定理计算似然比。基于 6.2 节中的基本考虑，本节假设待检测信号和噪声均为高斯随机过程。对于非高斯信号的情况，在 6.4 节中有简要描述。本节首先讨论加性高斯噪声中的确知信号和部分参数确知信号检测问题，然后分析高斯噪声中的随机信号检测问题。

6.3.1 相干检测

实际应用中的很多信号检测问题都可以归结为下述假设检验模型：

$$\begin{aligned} H_0 &: Y_t = N_t + s_t^0, \quad 0 \leqslant t \leqslant T \\ H_1 &: Y_t = N_t + s_t^1, \quad 0 \leqslant t \leqslant T \end{aligned} \tag{6.3.1}$$

式中，$\{s_t^0;\ t\in[0,\ T]\}$和$\{s_t^1;\ t\in[0,\ T]\}$是两个完全已知的信号波形，$\{N_t;\ t\in[0,\ T]\}$表示加性噪声的随机过程。第 3 章中讨论的连续时间相干检测问题，在数字通信、雷达等诸多应用中均会涉及。为了研究式(6.3.1)描述的假设检验问题，可以考虑下面的等价模型

$$\begin{aligned} H_0 &: Y_t = N_t, \qquad\qquad 0 \leqslant t \leqslant T \\ H_1 &: Y_t = N_t + s_t, \quad 0 \leqslant t \leqslant T \end{aligned} \tag{6.3.2}$$

从观测值中消除信号$\{s_t^0;\ t\in[0,\ T]\}$，并令 s_t 表示式(6.3.1)中的信号差 $s_t^1-s_t^0$，就可以得到式(6.3.2)中的模型。假设噪声$\{N_t;\ t\in[0,\ T]\}$是零均值高斯随机过程且连续的自协方差函数为 $C_N(t,\ u)$，$(t,\ u)\in[0,\ T]^2$。进一步假设$\{s_t;\ t\in[0,\ T]\}$连续且可表示为

$$s_t = \sum_{k=1}^{\infty} \hat{s}_k \psi_k(t), \quad 0 \leqslant t \leqslant T \tag{6.3.3}$$

式中$\{\psi_k\}_{k=1}^{\infty}$是 C_N 的正交特征函数集，$\hat{s}_k$ 是信号在 ψ_k 上$\{s_t;\ 0\leqslant t\leqslant T\}$的分量，即

$$\hat{s}_k = \int_0^T s_t \psi_k(t)\,\mathrm{d}t, \quad k = 1,2,\cdots \tag{6.3.4}$$

下面解释假设信号形如式(6.3.4)的原因。对于该表达式不成立的情况，将在后面予以讨论。

在对噪声$\{N_t;\ t\in[0,\ T]\}$进行 Karhunen- Loéve 展开后，结合式(6.3.3)，式(6.3.2)可改写为

$$\begin{aligned} H_0 &: Y_t = \sum_{k=1}^{\infty} \hat{N}_k \psi_k(t), \qquad\qquad 0 \leqslant t \leqslant T \\ H_1 &: Y_t = \sum_{k=1}^{\infty} (\hat{N}_k + \hat{s}_k) \psi_k(t), \quad 0 \leqslant t \leqslant T \end{aligned} \tag{6.3.5}$$

其中

$$\hat{N}_k = \int_0^T \psi_k(t) N_t\,\mathrm{d}t, \quad k = 1,2,\cdots \tag{6.3.6}$$

从式(6.3.2)和式(6.3.4)～式(6.3.6)可以看出，在两个假设下均可以将观测过程表示为

$$Y_t = \sum_{k=1}^{\infty} Z_k \psi_k(t), \quad 0 \leqslant t \leqslant T \tag{6.3.7}$$

式中

$$Z_k = \int_0^T Y_t \psi_k(t)\,\mathrm{d}t = \begin{cases} \hat{N}_k, & H_0\ 假设 \\ \hat{N}_k + \hat{s}_k, & H_1\ 假设 \end{cases} \tag{6.3.8}$$

因此式(6.3.2)中的假设等价于

$$\begin{aligned} H_0: Z_k &= \hat{N}_k, \qquad k = 1,2,\cdots \\ H_1: Z_k &= \hat{N}_k + \hat{s}_k, \quad k = 1,2,\cdots \end{aligned} \tag{6.3.9}$$

由 6.2 节可知$\{\hat{N}_k\}_{k=1}^{\infty}$是服从$\hat{N}_k \sim \mathcal{N}(0, \lambda_k)$分布的独立随机变量序列，其中$\{\lambda_k\}_{k=1}^{\infty}$是$C_N$的特征值序列。由于$\{\hat{s}_k\}_{k=1}^{\infty}$是已知序列，截取$n$个观测值后，式(6.3.9)所描述的检测问题就等同于 3.2 节中高斯噪声下离散时间相干检测的问题。特别地，基于式(6.3.9)中的前n个观测值的似然比可以写为

$$f_n(Z) = \exp\left\{\sum_{k=1}^{n} \hat{s}_k Z_k/\lambda_k - \frac{1}{2}\sum_{k=1}^{n} \hat{s}_k^2/\lambda_k\right\} \tag{6.3.10}$$

式中Z表示$\{Z_k\}_{k=1}^{\infty}$。根据 Grenander 定理，当集合的似然比L有限时，序列$\{f_n(Z)\}$在P_0和P_1测度下分别依概率收敛于L；当L无限时，序列在P_1测度下依概率发散。研究式(6.3.10)的收敛性可以得到下面的结论。

命题 6.3.1 Grenander 二分法

令P_0和P_1分别表示式(6.3.9)所描述的假设检验问题在假设H_0和H_1下的概率测度，那么有下面结论：

1) 如果$\sum_{k=1}^{\infty}(\hat{s}_k)^2/\lambda_k = \infty$，那么$P_0 \perp P_1$；

2) 如果$\sum_{k=1}^{\infty}(\hat{s}_k)^2/\lambda_k < \infty$，那么$P_0 \equiv P_1$且

$$\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(Z) = \exp\left[\sum_{k=1}^{\infty} \hat{s}_k Z_k/\lambda_k - \frac{1}{2}\sum_{k=1}^{\infty}(\hat{s}_k)^2/\lambda_k\right] \tag{6.3.11}$$

其中指数项的和是均方和。

证明 为了证明性质 1)，假设$\sum_{k=1}^{\infty}(\hat{s}_k)^2/\lambda_k = \infty$，在$b>0$条件下考虑概率

$$P_1(f_n(Z) > b) = P_1\left(\sum_{k=1}^{n} \hat{s}_k Z_k/\lambda_k - \frac{d_n^2}{2} > \log b\right) \tag{6.3.12}$$

式中$d_n^2 \triangleq \sum_{k=1}^{n}(\hat{s}_k)^2/\lambda_k$。在$H_1$假设下，$Z_1$，$Z_2$，…，$Z_n$是服从$\mathcal{N}(\hat{s}_k, \lambda_k)$的独立随机变量，这意味着

$$\sum_{k=1}^{n} \hat{s}_k Z_k/\lambda_k - \frac{d_n^2}{2} \sim \mathcal{N}\left(\frac{1}{2}d_n^2, d_n^2\right) \tag{6.3.13}$$

进而有

$$P_1(f_n(Z) > b) = 1 - \Phi\left(\frac{\log b}{d_n} - \frac{d_n}{2}\right) \tag{6.3.14}$$

式中d_n是d_n^2的正平方根($d_n = \sqrt{d_n^2}$)。如果$\lim_{n\to\infty} d_n^2 = \infty$成立，那么

$$\lim_{n\to\infty} P_1(f_n(Z) > b) = 1 - \Phi(-\infty) = 1, \quad b > 0 \tag{6.3.15}$$

即$f_n(Z)$在P_1概率测度下发散且$L = \infty$的观测集以概率 1 满足H_1假设，这意味着$P_0 \perp P_1$。

为了证明性质 2)，需要证明：如果 $\sum_{k=1}^{\infty}(\hat{s}-k)^2/\lambda_k<\infty$ 成立，则 $P_1 \ll P_0$ 和 $P_0 \ll P_1$ 成立。根据 Grenander 定理，只要能够证明 H_1 假设下 $f_n(Z)$ 在概率意义上收敛就能证明 $P_1 \ll P_0$。等价地，需要证明 $\log f_n(Z)$ 在 H_1 下依概率收敛。为此，需要验证 H_1 假设下 $\log f_n(Z)$ 在均方意义上收敛，这也就意味着依概率收敛[参看 Thomas(1986)]。

对数似然比 $\log f_n(Z)$ 可写成

$$\log f_n(Z)=\sum_{k=1}^{n}\hat{s}_k Z_k/\lambda_k-\frac{d_n^2}{2}=\sum_{k=1}^{n}\hat{s}_k X_k/\lambda_k+\frac{d_n^2}{2} \tag{6.3.16}$$

式中 $X_k \triangleq Z_k-\hat{s}_k$。由于 $\lim\limits_{n\to\infty}d_n^2<\infty$，因此只需要证明式(6.3.16)右边第一项是否收敛。根据 Parzen(1962)，序列 $\{W_k\}_{k=1}^{\infty}$ 在均方意义下收敛的充要条件是满足柯西准则，即

$$\lim_{n\to\infty}\sup_{m\geqslant n}E\{(W_m-W_n)^2\}=0 \tag{6.3.17}$$

定义 $W_n \triangleq \sum\limits_{k=1}^{n}\hat{s}_k X_k/\lambda_k$ 后，有

$$E_1\{(W_m-W_n)^2\}=E_1\left\{\left(\sum_{k=n+1}^{m}\hat{s}_k X_k/\lambda_k\right)^2\right\}=\sum_{k=n+1}^{m}\hat{s}_k^2/\lambda_k \tag{6.3.18}$$

上式中第二个等式成立是因为在 H_1 下独立随机变量 X_1，X_2，…，X_n 服从 $\mathcal{N}(0,\lambda_k)$ 分布。由于式(6.3.18)的被加数非负，所以

$$\sup_{m\geqslant n}E_1\{(W_m-W_n)^2\}=\sum_{k=n+1}^{\infty}(\hat{s}_k)^2/\lambda_k$$

如果 d_n^2 收敛，则在 $n\to\infty$ 时上式趋于 0，即条件 $\lim\limits_{n\to\infty}d_n^2<\infty$ 是 H_1 假设下 $\log f_n(Z)$ 均方收敛的充分条件。换言之，$\lim\limits_{n\to\infty}d_n^2<\infty$ 是 $P_1 \ll P_0$ 的充分条件。

为了证明 $\lim d_n^2<\infty$ 是 $P_0 \ll P_1$ 的充分条件，可以将 P_0 和 P_1 互换，然后证明在 H_0 假设下 $1/f_n(Z)$ 依概率收敛。这和在 H_0 假设下证明 $-\log f_n(Z)$ 依概率收敛等价，因此省略相应证明过程。至此，完成了命题 6.3.1 的全部证明。 □

命题 6.3.1 不但指出 $P_1 \ll P_0$ 成立的条件，更重要的是，它表明式(6.3.9)所描述的假设检验问题的两个测度或者相等或者奇异[式(6.3.2)所描述的假设检验问题也与之类似]，决定相等还是奇异的关键参数是

$$d^2 \triangleq \sum_{k=1}^{\infty}(\hat{s}_k)^2/\lambda_k \tag{6.3.19}$$

根据 Mercer 定理，有

$$\sum_{k=1}^{\infty}\lambda_k=\int_0^T C_N(t,t)\mathrm{d}t=\int_0^T E\{N_t^2\}\mathrm{d}t<\infty \tag{6.3.20}$$

式中积分值有限是因为 C_N 连续。既然 $\lambda_k=\mathrm{var}(Z_k)\geqslant 0$，即 $\lim\limits_{k\to\infty}\lambda_k=0$，那么 $\sum\limits_{k=1}^{\infty}(\hat{s}_k)^2<\infty$ 是 d^2 有限的必要条件。根据式(6.3.3)，信号在$[0,T]$上的能量可表示为

$$\int_0^T s_t^2\mathrm{d}t=\sum_{k=1}^{\infty}(\hat{s}_k)^2 \tag{6.3.21}$$

因此，为了使式(6.3.19)中的问题非奇异，信号的能量必须有限。然而，有限的信号能量并不足以确保检测问题没有奇异性，因为 d^2 有限仅要求 $(\hat{s}_k)^2/\lambda_k$ 可和，这意味着信号与

噪声自相关的特征结构之间具有微妙关系。

假设 $d^2<\infty$，贝叶斯、极小化极大和尼曼-皮尔逊最佳检验的判决准则均可表示为

$$\delta(Y_0^T)=\begin{cases}1, & \sum_{k=1}^{\infty}\hat{s}_k Z_k/\lambda_k > \log\tau+\frac{1}{2}d^2 \triangleq \tau' \\ \gamma, & \sum_{k=1}^{\infty}\hat{s}_k Z_k/\lambda_k = \log\tau+\frac{1}{2}d^2 \triangleq \tau' \\ 0, & \sum_{k=1}^{\infty}\hat{s}_k Z_k/\lambda_k < \log\tau+\frac{1}{2}d^2 \triangleq \tau'\end{cases} \tag{6.3.22}$$

与之对应的检测器结构如图 6.3.1 所示。注意，该结构包含无限组连续时间相关器（或匹配滤波器），它们分别对应噪声的不同特征函数。正如下面所看到的，从应用出发，最佳检测器可以表示为更简便的形式，而图 6.3.1 的表示形式非常便于解释该问题。特别地，特征函数$\{\psi_k\}$可以认为是包含噪声功率的正交"方向"的集合，λ_k 是 ψ_k 方向的噪声功率，$\hat{s}_k$ 是该方向的信号部分。无论是 H_0 假设还是 H_1 假设，ψ_k 正交都使得在 ψ_k 方向上的相关器 $Z_k=\int_0^T\psi_k(t)Y_t\mathrm{d}t$ 的输出序列$\{Z_k\}_{k=1}^{\infty}$独立。检测器将这些独立的输出按照与 3.2 节中讨论的离散时间问题类似的方式组合起来。

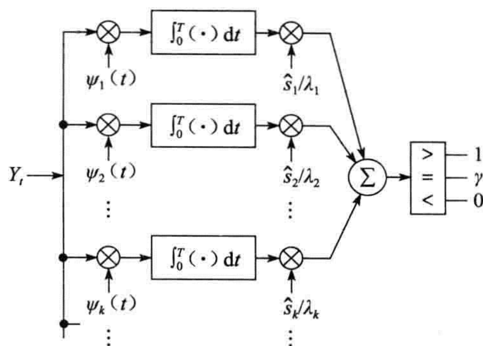


图 6.3.1 高斯噪声中的最优相干检测结构

在图 6.3.1 中，仅有信号时，输出 $\sum_{k=1}^{\infty}(\hat{s}_k)^2/\lambda_k=d^2$ 是确定量；仅有噪声时，输出 $\sum_{k=1}^{\infty}\hat{s}_k\hat{N}_k/\lambda_k$ 是零均值且方差为 d^2 的随机量。因此，图 6.3.1 中的判决门限可理解为信噪比，即

$$\mathrm{SNR}_0=\frac{(\text{输出信号的幅度})^2}{\text{输出噪声的功率}}=d^2 \tag{6.3.23}$$

可以看出，奇异性的条件可以理解为信噪比无穷大的情况，此时，可根据最优检测器的输出来计算信噪比。

统计量 $\sum_{k=1}^{\infty}\hat{s}_k Z_k/\lambda_k$ 是独立高斯随机变量$\{Z_k\}_{k=1}^{\infty}$的均方和，因此它也是高斯的。在 H_0 假设下，有 $Z_k\sim\mathcal{N}(0,\lambda_k)$，且有

$$\sum_{k=1}^{\infty}\frac{\hat{s}_k Z_k}{\lambda_k}\sim\mathcal{N}(0,d^2) \tag{6.3.24}$$

在 H_1 假设下，有 $Z_k\sim\mathcal{N}(\hat{s}_k,\lambda_k)$，即

$$\sum_{k=1}^{\infty}\frac{\hat{s}_k Z_k}{\lambda_k}\sim\mathcal{N}(d^2,d^2) \tag{6.3.25}$$

由式(6.3.24)和式(6.3.25)，可以计算式(6.3.22)中最优判决的性能。特别地，虚警概率由下式给出

$$P_0(\Gamma_1) = P_0\left(\sum_{k=1}^{\infty}\frac{\hat{s}_k Z_k}{\lambda_k} \geqslant \log\tau + \frac{d^2}{2}\right) = 1 - \Phi\left(\frac{\log\tau}{d} + \frac{d}{2}\right) \tag{6.3.26}$$

漏警概率由下式给出

$$P_1(\Gamma_0) = P_1\left(\sum_{k=1}^{\infty}\frac{\hat{s}_k Z_k}{\lambda_k} \geqslant \log\tau + \frac{d^2}{2}\right) = \Phi\left(\frac{\log\tau}{d} - \frac{d}{2}\right)$$

将上面的两个等式与式(2.2.31)中高斯误差下的标量位置检验性能对比可知，除了 d^2 的定义存在差异，虚警概率和漏警概率的计算式与标量模型完全相同。例如，与图 2.2.3 中所描述的一致，贝叶斯判决在均匀代价和相等先验分布下的贝叶斯风险为

$$r(\delta_B) = \frac{1}{2}P_0(\Gamma_1) + \frac{1}{2}P_1(\Gamma_0) = 1 - \Phi\left(\frac{d}{2}\right) \tag{6.3.27}$$

类似地，对虚警概率为 α 的尼曼-皮尔逊检测器，其判决门限 τ' 为 $\Phi^{-1}(1-\alpha)/d$，检测概率为

$$P_D = 1 - \Phi[\Phi^{-1}(1-\alpha) - d] \tag{6.3.28}$$

相应的接收机工作特性曲线见图 2.4.4。除了可能存在奇异性之外，求解出检测统计量之后，在高斯噪声背景下，连续时间信号的相干检测与离散时间信号的相干检测完全相同。

在给出似然比的另一种表示方式之前，首先简要讨论信号$\{s_t;\ t\in[0,\ T]\}$不能表示成式(6.3.3)的情况。如前所述，ψ_k 可以认为是一个含有噪声功率的正交“方向”的集合。因此，在与所有 ψ_k 都正交的方向上，是没有噪声功率的。如果信号能量有限，但却不能在噪声特征函数下表示，则信号在一个与所有的 ψ_k 都正交的方向上有能量。这也意味着存在某个方向，它上面只有信号能量而没有噪声能量，因此通过寻找此方向可以很好地实现奇异性检测。

特别地，考虑函数

$$f(t) \triangleq s_t - \sum_{k=1}^{\infty}\hat{s}_k\psi_k(t), \quad 0 \leqslant t \leqslant T$$

它与所有的 ψ_k 正交，因为对于任意的正整数 m，有

$$\int_0^T f(t)\psi_m(t)\,dt = \int_0^T s_t\psi_m(t)\,dt - \sum_{k=1}^{\infty}\hat{s}_k\int_0^T \psi_m(t)\psi_m(t)\,dt = \hat{s}_m - \hat{s}_m = 0$$

如果考虑判决统计量 $\int_0^T f(t)Y_t\,dt$，它对信号的响应为

$$\int_0^T f(t)s_t\,dt = \int_0^T s_t^2\,dt - \sum_{k=1}^{\infty}(\hat{s}_k)^2$$

表达式 $\sum \hat{s}_k\psi_k$ 不完备时上式非负。同样，它对噪声的响应为

$$\int_0^T f(t)N_t\,dt = \sum_{k=1}^{\infty}\hat{N}_k\int_0^T f(t)\psi_k(t)\,dt = 0$$

因此通过检测单个连续时间相关器的输出 $\int_0^T f(t)Y_t\,dt$，可以很好地判决信号是否存在。理解该问题的另一种方式是，$f(t)$是 C_N 的特征函数，记为 ψ_k，对应的特征值为 $\lambda_k=0$。此时，若$(\hat{s}_k)^2>0$，则 $d^2=\infty$，即导致奇异。

命题 6.3.1 给出的似然比表达式，对分析式(6.3.2)所描述的假设检测问题的性能和

结构非常有用。近似最优检测器的实现方法，即对图 6.3.1 所示的无限长滤波器组进行截断处理，使得信噪比损失 $\sum_{k=n+1}^{\infty}(\hat{s}_k)^2/\lambda_k$ 可以忽略。不过，如果似然比采用 Pitcher 定理给出的形式，检测器将具有更为有效的实现方式。下面证明 Pitcher 定理。

命题 6.3.2　Pitcher 定理

假设函数 $H:[0,T]\to\mathbf{R}$ 满足有界变分，

$$s_t = \int_0^T C_N(t,u)\mathrm{d}H(u),\quad 0\leqslant t\leqslant T \tag{6.3.29}$$

则 $P_0\equiv P_1$

$$\log\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(Y_0^T) = \int_0^T Y_t\mathrm{d}H(t) - \frac{1}{2}\int_0^T s_t\mathrm{d}H(t) \tag{6.3.30}$$

以及

$$d^2 = \int_0^T s_t\mathrm{d}H(t) \tag{6.3.31}$$

证明　假设$\{s_t;\ t\in[0,T]\}$可以表述为式(6.3.29)，根据 Mercer 定理，C_N 可表示成一致绝对收敛级数的和，即

$$C_N(t,u) = \sum_{k=1}^{\infty}\lambda_k\psi_k(t)\psi_k(u),\quad (t,u)\in[0,T]^2 \tag{6.3.32}$$

将上式代入式(6.3.29)，并交换积分和求和的顺序，可以得到

$$s_t = \sum_{k=1}^{\infty}\lambda_k\psi_k(t)\int_0^T\psi_k(u)\mathrm{d}H(u),\quad 0\leqslant t\leqslant T \tag{6.3.33}$$

将上式与式(6.3.3)相比较，可以看出

$$\int_0^T\psi_k(u)\mathrm{d}H(u) = \frac{\hat{s}_k}{\lambda_k},\quad k=1,2,\cdots \tag{6.3.34}$$

[注意，式(6.3.33)意味着式(6.3.3)有效]。考虑到 H 的变分有界且 C_N 的连续和在 $[0,T]^2$ 上有界，那么下式中的二重积分取有限值

$$\int_0^T\int_0^T C_N(t,u)\mathrm{d}H(u)\mathrm{d}H(t) \tag{6.3.35}$$

利用 Mercer 定理和式(6.3.34)，积分可改写为

$$\begin{aligned}\sum_{k=1}^{\infty}\lambda_k\int_0^T\int_0^T\psi_k(t)\psi_k(u)\mathrm{d}H(t)\mathrm{d}H(u) &= \sum_{k=1}^{\infty}\lambda_k\int_0^T\psi_k(t)\mathrm{d}H(t)\int_0^T\psi_k(u)\mathrm{d}H(u)\\ &= \sum_{k=1}^{\infty}\lambda_k\frac{\hat{s}_k}{\lambda_k}\frac{\hat{s}_k}{\lambda_k} = \sum_{k=1}^{\infty}\frac{(\hat{s}_k)^2}{\lambda_k} = d^2\end{aligned} \tag{6.3.36}$$

因此，$d^2<\infty$ 成立，且根据命题 6.3.1 有 $P_0\equiv P_1$ 成立。此外，因为 $\int_0^T s_t\mathrm{d}H(t)$ 等于式(6.3.35)的积分，根据式(6.3.36)可以直接推出式(6.3.31)成立。

由式(6.3.7)可知，观测信号可表示成 $Y_t=\sum_{k=1}^{\infty}Z_k\psi_n(t),0\leqslant t\leqslant T$ 的形式。再次利用式(6.3.34)，可以得到

$$\int_0^T Y_t\mathrm{d}H(t) = \sum_{k=1}^{\infty}Z_k\int_0^T\psi_k(t)\mathrm{d}H(t) = \sum_{k=1}^{\infty}\frac{\hat{s}_k Z_k}{\lambda_k} \tag{6.3.37}$$

综合式(6.3.37)、式(6.3.11)和式(6.3.31)后，可推出式(6.3.30)成立。至此，Pitcher 定理证明完毕。 □

相比式(6.3.2)而言，Pitcher 定理给出了最优检测器的更为简洁的实现形式。尤其是式(6.3.30)仅需计算单个积分 $\int_0^T Y_t \mathrm{d}H(t)$，而非求解积分序列 $\int_0^T \psi_k(t) Y_t \mathrm{d}t$。同样，求解 H 只涉及一个积分等式，即式(6.3.29)(也称 Pitcher 等式)，而非求解所有的特征函数等式 $\lambda\psi(t) = \int_0^T C_N(t,u)\psi(u)\mathrm{d}u, 0 \leqslant t \leqslant T$。

假设 $H(t)$ 可微，即 $h(t) = \mathrm{d}H(t)/\mathrm{d}t$，Pitcher 定理中的似然比与离散时间模型的似然比之间的关系非常明显。此时，积分式(6.3.29)可以写为

$$s_t = \int_0^T C_N(t,u)h(u)\mathrm{d}u, \quad 0 \leqslant t \leqslant T \tag{6.3.38}$$

最优检测统计量变为

$$\log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(Y_0^T) = \int_0^T Y_t h(t)\mathrm{d}t - \frac{1}{2}\int_0^T s_t h(t)\mathrm{d}t \tag{6.3.39}$$

有趣的是，将上面两式分别与式(3.2.25)和式(3.2.23)比较，用 $\boldsymbol{h}$ 替换 $\tilde{\boldsymbol{s}}$。可以看到式(6.3.38)是向量 $\boldsymbol{s} = \boldsymbol{\Sigma}_Y \boldsymbol{h}$ 在连续时间下的等价表达式，或者

$$s_k = \sum_{j=1}^{n} (\boldsymbol{\Sigma}_Y)_{kj} h_j, \quad 1 \leqslant k \leqslant n \tag{6.3.40}$$

显然，上式给出了离散信号最优相干检测中“伪信号”$\boldsymbol{h}$ 的形式。类似地，式(6.3.39)的离散形式为

$$\log \frac{p_1(\boldsymbol{Y})}{p_0(\boldsymbol{Y})} = \boldsymbol{h}^{\mathrm{T}}\boldsymbol{T} - \frac{1}{2}\boldsymbol{h}^{\mathrm{T}}\boldsymbol{s} = \sum_{k=1}^{n} h_k Y_k - \frac{1}{2}\sum_{k=1}^{n} h_k s_k \tag{6.3.41}$$

综上，Pitcher 定理给出了连续时间模型的似然比的直接表述方式，而定理 6.3.1 描述了预白化后的似然比。

上述解释能够很好地将 Pitcher 定理与相应的离散结果有机联系起来。不过，对于连续的 C_N，通常情况下 $H(t)$ 是不可微的。因此，有必要进一步给出式(6.3.30)和式(6.3.31)的更一般性解释。下面用具体例子来说明如何应用 Pitcher 定理。

例 6.3.1 高斯白噪声下的相干检测

高斯白噪声是实际应用中最普遍的噪声模型。在离散采样下，高斯白噪声是独立同分布且均值为零的宽平稳随机变量序列，其自协方差矩阵可表示为

$$C_N(k,l) = C_N(k-l,0) \triangleq C_N(k-l) = \frac{N_0}{2}\delta_{k,l}, \quad k,l \in \mathbf{Z} \tag{6.3.42}$$

其中 $\delta_{k,l}$ 是克罗内克冲激函数($k \neq l$ 时 $\delta_{k,l} = 0$；$k = l$ 时 $\delta_{k,l} = 1$)，$N_0/2$ 是序列中每一个元素的方差。正如第 5 章所述，之所以称为“白噪声”是因为它的能量谱密度为常数，即

$$\phi_N(\omega) \triangleq \sum_{k=-\infty}^{\infty} C_N(k)\mathrm{e}^{-\mathrm{i}\omega k} = \frac{N_0}{2}, \quad -\pi \leqslant \omega \leqslant \pi \tag{6.3.43}$$

连续时间下高斯白噪声的自协方差函数可以等价表示为

$$C_N(t,u) = C_N(t-u,0) \triangleq C_N(t-u) = \frac{N_0}{2}\delta(t-u), \quad t,u \in \mathbf{R} \tag{6.3.44}$$

其中 δ 是狄拉克冲激函数[⊖]。连续时间信号的功率谱密度对应式(6.3.43)的傅里叶积分，即

$$\phi_N(\omega)=\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}\omega\tau}\frac{N_0}{2}\delta(\tau)\mathrm{d}t=\frac{N_0}{2},\quad -\infty<\omega<\infty \tag{6.3.45}$$

因此，与式(6.3.44)所描述的自协方差函数对应的随机过程具有平坦的功率谱密度，$N_0/2$称为噪声的功率谱幅度。

遗憾的是，真实的物理现象和数学意义上的随机过程均不存在理想的高斯白噪声。首先，对于任何一个物理现象都不可能在所有的频率分量中有相等的非零能量。其次，尽管实际中不存在理想高斯白噪声的原因较为复杂[参见 Skorohod(1974)]，但是从二阶随机过程来看，由于 $E\{N_t^2\}=C_N(t,t)=(N_0/2)\delta(t-t)=\infty$，所以确实不存在理想的高斯白噪声。不过，式(6.3.44)给出的自协方差结构在分析建模中非常有用，且常用来对大带宽信号进行建模。一般而言，分析线性系统在白噪声下的输出特性相对容易。在信号检测方面，可以按照下面的步骤来严格处理白噪声问题。

假设信号检测问题可描述为

$$\begin{aligned}H_0&:Y_t=N_t, &0\leqslant t\leqslant T\\ H_1&:Y_t=N_t+s_t, &0\leqslant t\leqslant T\end{aligned} \tag{6.3.46}$$

其中，$\{N_t;\ t\in[0,\ T]\}$表示高斯白噪声，$\{s_t;\ t\in[0,\ T]\}$是确知信号。先不考虑其他问题，由于$\{N_t;\ t\in[0,\ T]\}$没有二阶项，不能利用本节的内容直接推出其似然比。然而，可以构造一个等价的问题，进一步通过对式(6.3.46)进行积分求解，得到下面的信号检测模型

$$\begin{aligned}H_0&:X_t=W_t, &0\leqslant t\leqslant T\\ H_1&:X_t=W_t+m_t, &0\leqslant t\leqslant T\end{aligned} \tag{6.3.47}$$

其中，$X_t\triangleq\int_0^t Y_u\mathrm{d}u, W_t\triangleq\int_0^t N_u\mathrm{d}u, m_t\triangleq\int_0^t s_u\mathrm{d}u$。容易证明：$\{W_t;\ t\in[0,\ T]\}$是零均值高斯过程，其自协方差矩阵为

$$C_W(t,u)=\frac{N_0}{2}\min\{t,u\},\quad (t,u)\in[0,T]^2 \tag{6.3.48}$$

拥有上述特性的随机过程称为维纳过程。由于 $C_W(t,\ u)$在$[0,\ T]^2$上连续，所以可根据之前的方法计算似然比。具体步骤是：首先求解式(6.3.47)所描述的假设检验对的 Pitcher 等式(6.3.29)，即求解

$$m_t=\int_0^T C_W(t,u)\mathrm{d}H(u),\quad 0\leqslant t\leqslant T \tag{6.3.49}$$

代入 m_t 和 C_W 后，得到

$$\int_0^T s_u\mathrm{d}u=\frac{N_0}{2}\int_0^T\min\{t,u\}\mathrm{d}H(u)$$

⊖ 狄拉克冲激函数是广义函数，其定义为在 $x=0$ 处连续且满足$\int_{-\infty}^{\infty}f(x)\delta(x)\mathrm{d}x=f(0)$ 的任意函数。实际上，没有一个真实的函数能具有这样的特性，因为它要求 δ 满足 $\delta(0)=+\infty$ 和$\lim\limits_{x\to 0}\delta(x)=0$。

$$= \frac{N_0}{2}\int_0^t u\mathrm{d}H(u) + \frac{N_0}{2}t\int_0^T \mathrm{d}H(u)$$

$$= \frac{N_0}{2}\int_0^t u\mathrm{d}H(u) + \frac{N_0}{2}t[H(T) - H(t)] \tag{6.3.50}$$

对等式右边的第一项进行分部积分后，有

$$\int_0^t s_u \mathrm{d}u = \frac{N_0}{2}[tH(t) - \int_0^t H(u)\mathrm{d}u] + \frac{N_0}{2}t[H(T) - H(t)]$$

$$= \frac{N_0}{2}\{\int_0^t [H(T) - H(u)]\mathrm{d}u\}, \quad 0 \leqslant t \leqslant T \tag{6.3.51}$$

假设信号在[0，T]上有界，Pitcher 等式的解为

$$H(t) = \begin{cases} H(T) - \dfrac{2}{N_0}s_t, & 0 \leqslant t \leqslant T \\ H(T), & t = T \end{cases} \tag{6.3.52}$$

其中 $H(T)$是任意的，函数 s_t 如图 6.3.2 所示。

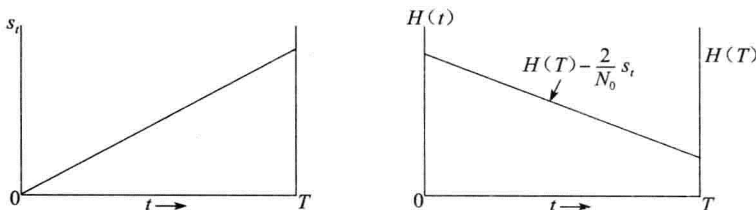


图 6.3.2 维纳噪声的 Pitcher 等式求解示意图

进一步假设信号在终点 $t=T$ 处连续，因为 H 在 $t=T$ 处有 $2s_t/N_0$ 的跳跃，那么似然比为

$$\log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(X_0^T) = -\frac{2}{N_0}\int_0^T \left(X_t - \frac{1}{2}m_t\right)\mathrm{d}s_t + \frac{2}{N_0}\left(X_T - \frac{1}{2}m_T\right)s_T \tag{6.3.53}$$

又因为 $s_t = \dfrac{\mathrm{d}m_t}{\mathrm{d}t}$，分部积分后可以得到

$$\log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(X_0^T) = \frac{2}{N_0}\int_0^T s_t \mathrm{d}\left(X_t - \frac{1}{2}m_t\right)$$

$$= \frac{2}{N_0}\int_0^T s_t \mathrm{d}X_t - \frac{1}{N_0}\int_0^T s_t \mathrm{d}m_t$$

$$= \frac{2}{N_0}\int_0^T s_t \mathrm{d}X_t - \frac{1}{N_0}\int_0^T s_t^2 \mathrm{d}t \tag{6.3.54}$$

上式就是著名的 Cameron-Martin 公式。下面将要讨论，对于满足 $\int_0^T s_t^2 \mathrm{d}t < \infty$ 的任意信号，Cameron-Martin 公式都成立。如果将初始观测值 Y_t 视为 X_t 的微分(因为初始信号包含白噪声，这仅仅是启发式的解释)，那么上式的物理意义将非常明了。进一步将式(6.3.54)写为

$$\log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(Y_0^T) = \frac{2}{N_0}\int_0^T s_t Y_t \mathrm{d}t - \frac{1}{N_0}\int_0^T s_t^2 \mathrm{d}t \tag{6.3.55}$$

在离散时间模型中，利用包含服从$\mathcal{N}(0, N_0/2)$的高斯加性白噪声的观测值 Y_1，…，Y_n

检测确知信号 s_1，…，s_n 的似然比[见式(3.2.9)]为

$$\log\frac{p_1(Y_0^n)}{p_0(Y_0^n)}=\frac{2}{N_0}\sum_{k=1}^{n}s_kY_k-\frac{1}{N_0}\sum_{k=1}^{n}s_k^2 \tag{6.3.56}$$

显然，式(6.3.55)和式(6.3.56)类似。特别地，式(6.3.55)是相关器或匹配滤波器的连续时间形式，相应的检测判决是

$$\delta(Y_0^T)=\begin{cases}1, & \dfrac{\mathrm{d}P_1}{\mathrm{d}P_0}(Y_0^T)>\tau\\ \gamma, & \dfrac{\mathrm{d}P_1}{\mathrm{d}P_0}(Y_0^T)=\tau\\ 0, & \dfrac{\mathrm{d}P_1}{\mathrm{d}P_0}(Y_0^T)<\tau\end{cases} \tag{6.3.57}$$

具体检测流程如图 6.3.3 所示。

式(6.3.57)所描述的检测器的性能由 d^2 的数值决定，它可根据 Pitcher 定理按照下式计算

$$d^2=\int_0^T m_t\mathrm{d}H(t)=\frac{2}{N_0}\int_0^T m_t\mathrm{d}s_t+\frac{2}{N_0}m_Ts_T$$
$$=\frac{2}{N_0}\int_0^T s_t^2\mathrm{d}t \tag{6.3.58}$$

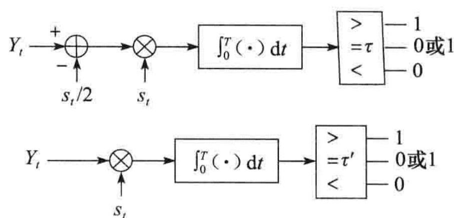


图 6.3.3 高斯白噪声下相干信号检测结构

显然，检测性能由总信号能量与噪声的功率谱幅度的比值决定。此外，由于信号能量有限，所以不会出现奇异性。

事实上，根据 Pitcher 定理，图 6.3.3 中接收器的结构可由白噪声模型(6.3.46)直接得到。假设 H 是可微的且记 $h(t)=\mathrm{d}H(t)/\mathrm{d}t$ 后，可以得到

$$s_t=\int_0^T C_N(t,u)h(u)\mathrm{d}u=\frac{N_0}{2}\int_0^T\delta(t-u)h(u)\mathrm{d}u=\frac{N_0}{2}h(t) \tag{6.3.59}$$

其中 $h(t)=(2/N_0)s_t$。相应的似然比可写成

$$\log\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(Y_0^T)=\frac{2}{N_0}\left[\int_0^T s_tY_t\mathrm{d}t-\frac{1}{2}\int_0^T s_t^2\mathrm{d}t\right] \tag{6.3.60}$$

这与式(6.3.55)相同。当然，不能这样直接应用 Pitcher 定理，但可以理解为噪声在较宽频带内具有平坦谱结构下的近似最优接收器结构。 ◀

针对其他不同结构的协方差矩阵模型，同样能得到相应的 Pitcher 等式的解，在本章习题中也会涉及一些具体求解实例。一般情况下，解 H 在观测期间可微，但在端点处不连续(如例 6.3.1 中维纳过程的情况)。由于这个原因，非常容易理解 $H(t)$ 是函数 $h(t)$ 的积分，且 $h(t)$ 在(0，T)内是实值，并且在端点处有奇异点(即狄拉克冲激函数)。基于上述理解，Kailath 在 1966 年给出了 Pitcher 方程的求解方法。

现在讨论关于 Pitcher 方程的最后一个问题，通常假设 Pitcher 方程具有如下形式：

$$C_N(t,u)=C_C(t,u)+\frac{N_0}{2}\delta(t-u),\quad (t,u)\in[0,T]^2 \tag{6.3.61}$$

其中 C_C 是连续函数，$N_0>0$，即噪声由连续分量和加性白化分量[⊖]构成。与例 6.3.1 相

⊖ 白化分量有时称为噪声电平。

同，根据 Pitcher 定理，式(6.3.52)的解通常是可微的，且导数 h 满足

$$s_t = \int_0^T C_C(t,u)h(u)\mathrm{d}u + \frac{N_0}{2}h(t), \quad 0 \leqslant t \leqslant T \tag{6.3.62}$$

该式也称为第二类 Fredholm 积分等式，与之对应的求解方法是众所周知的[参见 Lovitt (1950)]。$N_0=0$ 对应第一类 Fredholm 等式，它的解在 $t=0$ 或 $t=T$ 处有奇异点。加入 $N_0>0$ 的白噪声后可抑制这些奇异点。与例 6.3.1 一样，上述分析方法在一定程度上是非常缜密的。

6.3.2 参数未知的信号检测

与第 3 章的讨论一样，信号形式已知但包含某些未知参数是信号检测的常见模型。例如，对正弦信号而言，相位、频率及幅度等先验信息是未知的。

为研究参数未知的连续时间检测问题，考虑以下假设检验

$$\begin{aligned} H_0: Y_t &= N_t, & 0 \leqslant t \leqslant T \\ H_1: Y_t &= s_t(\theta) + N_t, & 0 \leqslant t \leqslant T \end{aligned} \tag{6.3.63}$$

其中，$s_t(\theta)$是在集合 Λ 上取值的未知参数 θ 的已知函数，$\{N_t;\ t\in[0,\ T]\}$表示功率谱幅度为 $N_0/2$ 的高斯白噪声。与例 6.3.1 一样，通过对观测值进行积分后，可得到更利于处理的形式[⊖]

$$\begin{aligned} H_0: X_t &= W_t, & 0 \leqslant t \leqslant T \\ H_1: X_t &= W_t + \int_0^t s_u(\theta)\mathrm{d}u, & 0 \leqslant t \leqslant T \end{aligned} \tag{6.3.64}$$

其中，$X_t = \int_0^t Y_u \mathrm{d}u$ 和$\{W_t; t\in[0,T]\}$是自协方差为 $C_W(t,u)=(N_0/2)\min\{t,u\}$ 的维纳过程。进一步假设$\int_0^T s_t^2(\theta)\mathrm{d}t<\infty$ 对所有 $\theta\in\Lambda$ 成立，那么对于给定的 θ，根据式(6.3.54)所描述的 Camron-Martin 公式，可得出似然比为

$$L_\theta(X_0^T) \triangleq \frac{\mathrm{d}P_\theta}{\mathrm{d}P_0}(X_0^T) = \exp\left\{\frac{2}{N_0}\int_0^T s_t(\theta)\mathrm{d}X_t - \frac{1}{N_0}\int_0^T s_t^2(\theta)\mathrm{d}t\right\} \tag{6.3.65}$$

假设参数集 $\Lambda\in\mathbf{R}^m$，根据之前讨论的结果，可以采用三种方法进行 H_1 与 H_0的检验。它们是：判决准则为

$$\max_\delta P_D(\delta|\theta), \quad \theta\in\Lambda, \quad \text{s.t.} \quad P_F(\delta)\leqslant\alpha \tag{6.3.66}$$

的 UMP 检验；在 H_1 下，在参数集 Λ 上指定先验分布 $w(\theta)$的检验(包括贝叶斯假设检验、极小化极大假设检验和尼曼-皮尔逊假设检验)；将广义似然比

$$\max_{\theta\in\Lambda} L_\theta(X_0^T) \tag{6.3.67}$$

与门限进行比较的最大似然比检验。

对于虚警概率为 α 的 UMP 检验，当且仅当可独立于 θ 选择关键域 $\Gamma_\theta=\{L_\theta(Y_0^T)>\tau_\theta\}$ 时，$P_0(\Gamma_\theta)=\alpha$ 对所有 $\theta\in\Lambda$ 才存在。因此，所有简单检验都可以通过比较平均似然比

$$\int_\Lambda L_\theta(X_0^T)w(\theta)\mathrm{d}\theta \tag{6.3.68}$$

与门限的大小来实现判决目的。

⊖ 6.4 节将会讨论一类广义高斯噪声过程可线性退化为维纳过程，即式(6.3.64)可视为加性高斯噪声的广义模型。

如果不进一步限制信号模型，那么就难以分析不同判决准则的异同。不过，如果给定 L_θ 具有式(6.3.65)所描述的形式，连续时间模型和离散时间模型的检测器就没有本质差别。下面的例子将说明这一问题，这个例子与离散域的例 3.2.5 比较相似。

例 6.3.2　白噪声中部分参数已知的正弦信号检测

考虑以下信号模型

$$s_t(\theta) = a_t \sin(\omega_c t + \theta), \quad 0 \leqslant t \leqslant T \tag{6.3.69}$$

其中，ω_c 是满足 $\omega_c T/\pi$ 为整数的已知频率；θ 为未知相位；$\{a_t;\ 0 \leqslant t \leqslant T\}$ 为已知波形，且满足

$$\frac{1}{T}\int_0^T a_t^2 \mathrm{d}t \triangleq \overline{a^2} < \infty$$

为简单起见，假设

$$\int_0^T a_t^2 \cos(2\omega_c t + \phi)\mathrm{d}t = 0 \tag{6.3.70}$$

对所有 $0 \leqslant \phi \leqslant 2\pi$ 成立。事实上，当 a_t^2 为常数或升余弦时，式(6.3.70)一定成立。此外，如果 a_t^2 的变化相比频率为 $2\omega_c$ 的正弦信号而言要慢得多，式(6.3.70)也近似成立。

给定 θ 后，式(6.3.64)的似然比为

$$L_\theta(X_0^T) = \exp\left\{\frac{2}{N_0}\int_0^T a_t \sin(\omega_c t + \theta)\mathrm{d}X_t - \frac{1}{N_0}\int_0^T a_t^2 \sin^2(\omega_c t + \theta)\mathrm{d}t\right\} \tag{6.3.71}$$

定义 Y_c 与 Y_s 为

$$Y_c = \int_0^T a_t \cos(\omega_c t)\mathrm{d}X_t$$

和

$$Y_s = \int_0^T a_t \sin(\omega_c t)\mathrm{d}X_t \tag{6.3.72}$$

利用三角等式 $\sin(a+b) = \cos a \sin b + \sin a \cos b$，式(6.3.71)的指数项中的第一项可变为

$$\frac{2}{N_0}(Y_c \sin\theta + Y_s \cos\theta)$$

利用三角等式 $2\sin^2(a) = 1 - \cos(2a)$，并代入式(6.3.70)后，指数项中的第二项可变为

$$\frac{\overline{a^2}T}{2N_0} - \frac{1}{2N_0}\int_0^T a_t^2 \cos(2\omega_c t + 2\theta)\mathrm{d}t = \frac{\overline{a^2}T}{2N_0} \tag{6.3.73}$$

因此，可得出

$$L_\theta(X_0^T) = \mathrm{e}^{\frac{-\overline{a^2}T}{2N_0}} \times \exp\left\{\frac{2}{N_0}(Y_c \sin\theta + Y_s \cos\theta)\right\} \tag{6.3.74}$$

将式(6.3.74)与例 3.2.5 所得的似然比进行比较，可看出除了 Y_c 与 Y_s 的定义不同外，离散时间与连续时间的条件似然比一致($T=n$)。因此，就例 3.2.5 及第 3 章习题中关于$\{a_t\}$与 θ 的多种假设模型所得的似然比而言，只要 Y_c 和 Y_s 采用式(6.3.72)的定义，在与之对应的连续时间模型中都成立。此外，由例 3.2.5 所推导出的性能表达式在连续时间情况下也成立。为说明该问题，注意到对给定的 θ，$\{X_t;\ 0 \leqslant t \leqslant T\}$ 在两种假设下均为高斯随机过程。又因为 Y_c 与 Y_s 是 X_0^T 的线性变换，因此它们也是高斯的。可见[⊖]，对[0，T]区

⊖ 该性质见 6.4 节。

间上的任意两个均方可积函数 f 和 g，有

$$E\left\{\int_0^T f(t)\mathrm{d}W_t\right\}=0 \tag{6.3.75}$$

和

$$E\left\{\int_0^T f(t)\mathrm{d}W_t\int_0^T g(t)\mathrm{d}W_t\right\}=\frac{N_0}{2}\int_0^T f(t)g(t)\mathrm{d}t \tag{6.3.76}$$

其中，$\{W_t;\ 0\leqslant t\leqslant T\}$ 是均值为零、协方差为 $(N_0/2)\min\{t,\ u\}$ 的维纳过程。进而，对给定的 θ，可直接得出

$$\begin{aligned}E\{Y_c\mid H_j\}&=j\frac{\overline{a^2}}{2}\sin\theta\\E\{Y_s\mid H_j\}&=j\frac{\overline{a^2}}{2}\cos\theta\\\operatorname{var}\{Y_c\mid H_j\}&=\operatorname{var}\{Y_s\mid H_j\}=\frac{N_0T}{4}\\\operatorname{cov}\{Y_c,Y_s\mid H_j\}&=0\end{aligned} \tag{6.3.77}$$

其中 $j=0$，1。将式(6.3.77)与例 3.2.5 对应的量进行比较，给定 θ 后，Y_c 与 Y_s 的统计特性与其离散时间对应的量是一致的。又因为似然比表达式也一致，所以例 3.2.5 的性能表达式对连续时间情况也成立。◀

可见，就相干检测而言，只要确定了对应的似然比，具有随机参数的连续时间信号检测与离散时间检测类似。式(6.3.64)与信号参数估计的联系将在第 7 章具体介绍。

6.4　高斯噪声中的随机信号检测

在 6.3 节中，已经给出了确知信号和部分参数未知信号的检测方法。不过，与第 3 章类似，在很多实际应用中需要将感兴趣的信号建模成完全随机过程。例如，在声纳和射电天文学里，传播媒介的随机扰动会破坏信号的参数化结构。本节将讨论高斯噪声中的随机信号检测问题。

特别地，考虑在谱幅度为 $N_0/2$ 的加性高斯白噪声中检测随机信号 $\{S_t;\ 0\leqslant t\leqslant T\}$。与前面处理白噪声的过程一样，首先对观测值积分得到以下模型

$$\begin{aligned}H_0:X_t&=W_t, && 0\leqslant t\leqslant T\\H_1:X_t&=\int_0^t S_u\mathrm{d}u+W_t, && 0\leqslant t\leqslant T\end{aligned} \tag{6.4.1}$$

式中，$\{W_t;\ t\in[0,\ T]\}$ 是维纳过程，$\int_0^t S_u\mathrm{d}u$ 表示 $\{S_u;\ 0\leqslant t\leqslant T\}$ 的均方积分。可见，模型中维纳噪声的约束条件并不苛刻，因为许多高斯噪声经过“白化处理”后都能线性转化成维纳噪声。本节主要讨论 $\{S_t;\ 0\leqslant t\leqslant T\}$（也是高斯随机过程）的情况，非高斯过程也将简要介绍。

6.4.1　维纳过程的初步结论

在讨论具体的信号检测方法之前，为便于理解，首先给出维纳过程的一些性质。在整个讨论过程中，$\{W_t;\ 0\leqslant t\leqslant T\}$ 表示参数为 $N_0/2$ 的维纳过程，即 $\{W_t;\ 0\leqslant t\leqslant T\}$ 是均值

为零、协方差为 $E\{W_tW_u\}=(N_0/2)\min\{t, u\}$的高斯过程。现将维纳过程的一些相关性质总结如下。

命题 6.4.1 维纳过程的性质

维纳过程$\{W_t;\ 0\leqslant t\leqslant T\}$具有如下性质：

1) $W_0=0$ 以概率 1 成立；

2) $E\{(W_t-W_s)(W_u-W_v)\}=(N_0/2)l\{(u, v)\cap(t, s)\}$对所有 $t>s\geqslant 0$ 及 $u>v\geqslant 0$ 成立，其中 $l\{(a, b)\}=b-a$；

3) 当且仅当 f 在$[0, T]$内平方可积，$\int_0^T f_t \mathrm{d}W_t$ 以均方积分形式存在；

4) 如果 $\int_0^T f_t \mathrm{d}W_t$ 存在，那么 $E\left\{\int_0^T f_t \mathrm{d}W_t\right\}=0$；

5) 如果 f 和 g 在$[0, T]$内平方可积，那么

$$E\left\{\int_0^T f_t \mathrm{d}W_t \int_0^T g_t \mathrm{d}W_t\right\}=\frac{N_0}{2}\int_0^T f_t g_t \mathrm{d}t$$

6) 如果 $\int_0^T f_t \mathrm{d}W_t$ 存在，那么当 $0\leqslant t\leqslant T$ 时，有

$$E\left\{W_t \int_0^T f_u \mathrm{d}W_u\right\}=\frac{N_0}{2}\int_0^t f_u \mathrm{d}u$$

证明 性质 1)和 2)可由 W_0^T 的自协方差结构直接得到。性质 3)可以由性质 2)和均方收敛的柯西准则[见式(6.3.17)]推出。根据 W_0^T 的二阶统计特性，假设可交换期望与均方极限的顺序后，根据第 5 章的内容易知性质 4)和性质 6)成立。接下来证明性质 5)，其他性质的证明留作练习。

为了证明性质 5)，首先，如果$\{X_n\}_{n=1}^{\infty}$与$\{Y_n\}_{n=1}^{\infty}$分别是在均方意义上收敛到 X 与 Y 的二阶序列，那么 $E\{X_nY_n\}\to E\{XY\}$。此外，$\int_0^T f_t \mathrm{d}W_t$ 是 $\sum_{i=1}^{n} f_{\xi_i^{(n)}}[W_{t_i^{(n)}}-W_{t_{i-1}^{(n)}}]$ 的均方极限，其中，$0=t_0^{(n)}<\cdots<t_1^{(n)}=T$，$\xi_i^{(n)}\in[t_{t-1}^{(n)}, t_i^{(n)}]$，$\Delta_n\triangleq\max\limits_{1\leqslant i\leqslant n}|t_i^{(n)}-t_{i-1}^{(n)}|$趋近 0。用 $g_{\xi_i^{(n)}}$ 代替 $f_{\xi_i^{(n)}}$ 后，$\int_0^T g_t \mathrm{d}W_t$ 可用同样的极限表示。因此可得到

$$\begin{aligned}E\left\{\int_0^T f_t \mathrm{d}W_t \int_0^T g_t \mathrm{d}W_t\right\}&=\lim_{\Delta_n\to 0} E\left\{\sum_{i=1}^{n} f_{\xi_i^{(n)}}[W_{t_i^{(n)}}-W_{t_{i-1}^{(n)}}]\sum_{j=1}^{n} g_{\xi_j^{(n)}}[W_{t_j^{(n)}}-W_{t_{j-1}^{(n)}}]\right\}\\&=\lim_{\Delta_n\to 0}\sum_{i=1}^{n}\sum_{j=1}^{n} f_{\xi_j^{(n)}} g_{\xi_i^{(n)}} E\{[W_{t_i^{(n)}}-W_{t_{i-1}^{(n)}}][W_{t_j^{(n)}}-W_{t_{j-1}^{(n)}}]\}\end{aligned} \tag{6.4.2}$$

由性质 2)可得

$$E\{[W_{t_i^{(n)}}-W_{t_{i-1}^{(n)}}][W_{t_j^{(n)}}-W_{t_{j-1}^{(n)}}]\}=\begin{cases}\dfrac{N_0}{2}[t_i^{(n)}-t_{i-1}^{(n)}], & i=j\\ 0, & i\neq j\end{cases} \tag{6.4.3}$$

将式(6.4.3)代入式(6.4.2)，可得

$$E\{\int_0^T f_t \mathrm{d}W_t \int_0^T g_t \mathrm{d}W_t\}=\lim_{\Delta_n\to 0}\sum_{i=1}^{n} f_{\xi_i^{(n)}} g_{\xi_i^{(n)}}[t_i^{(n)}-t_{i-1}^{(n)}]=\frac{N_0}{2}\int_0^t f_t g_t \mathrm{d}t \tag{6.4.4}$$

这就是需要证明的结果。 □

根据命题 6.4.1 中的性质，在命题 6.4.2 中可得到维纳滤波的另一种非常有用的等价表达形式。接下来，先定义正交函数的完备性。

定义 6.4.1　正交函数的完备性

记$[0, T]$上的正交函数集合为$\{\phi_k\}_{k=1}^{\infty}$，如果定义在$[0, T]$上的任意平方可积函数 f 都可以表示成如下形式

$$f_t = \sum_{k=1}^{\infty} \hat{f}_k \phi_k(t), \quad 0 \leqslant t \leqslant T \tag{6.4.5}$$

则称正交函数集$\{\phi_k\}_{k=1}^{\infty}$在$[0, T]$上**完备**。其中，$\hat{f}_k \triangleq \int_0^T f_t \phi_k(t) \mathrm{d}k, k = 1,2,\cdots$ 和式(6.4.5)在平方可积意义上收敛，即

$$\lim_{n\to\infty}\int_0^T \Big[f_t - \sum_{k=1}^{n} \hat{f}_k \phi_k(t) \Big]^2 \mathrm{d}t = 0 \tag{6.4.6}$$

命题 6.4.2　维纳过程的表示

假设$\{\phi_k\}_{k=1}^{\infty}$是$[0, T]$上的完备正交函数集，则维纳过程$\{W_t;\ 0 \leqslant t \leqslant T\}$可表示为

$$W_t = \sum_{k=1}^{\infty} \hat{W}_k \int_0^t \phi_k(u) \mathrm{d}u, \quad 0 \leqslant t \leqslant T \tag{6.4.7}$$

其中，$\hat{W}_k \triangleq \int_0^T \phi_k(t) \mathrm{d}W_t, k = 1,2,\cdots$，求和运算对应均方和。

证明　取 $t \in [0, T]$，有

$$E\Big\{\Big[W_t - \sum_{k=1}^{n} \hat{W}_k \int_0^t \phi_k(u) \mathrm{d}u \Big]^2\Big\} = E\{W_t^2\} - 2\sum_{k=1}^{n} E\{W_t \hat{W}_k\} \int_0^t \phi_k(u) \mathrm{d}u$$

$$+ \sum_{k=1}^{n} \sum_{l=1}^{n} E\{\hat{W}_k \hat{W}_l\} \int_0^t \phi_k(u) \mathrm{d}u \int_0^t \phi_l(u) \mathrm{d}u \tag{6.4.8}$$

因为 $E\{W_t^2\} = (N_0/2)t$，根据命题 6.4.1 可得

$$E\{W_t \hat{W}_k\} = \frac{N_0}{2} \int_0^t \phi_k(u) \mathrm{d}u$$

和

$$E\{\hat{W}_k \hat{W}_l\} = \frac{N_0}{2} \int_0^T \phi_k(t) \phi_l(t) \mathrm{d}t$$

又因为$\{\phi_k\}_{k=1}^{\infty}$是完备正交函数集，因此式(6.4.8)变为

$$\frac{N_0}{2}\Big[t - \sum_{k=1}^{n} \Big(\int_0^t \phi_k(u) \mathrm{d}u \Big)^2 \Big]$$

定义函数 $1_u^{(t)}$ 为

$$1_u^{(t)} = \begin{cases} 1, & 0 \leqslant u \leqslant t \\ 0, & t < u \leqslant T \end{cases} \tag{6.4.9}$$

并定义 $\hat{1}_k^{(t)} \triangleq \int_0^T 1_u^{(t)} \phi_k(u) \mathrm{d}u \equiv \int_0^t \phi_k(u) \mathrm{d}u, k = 1,2,\cdots$。由于$\int_0^T [1_u^t]^2 = t < \infty$，$\{\phi_k\}$ 完备则意味着

$$\lim_{n\to\infty}\int_0^T \Big[1_u^{(t)} - \sum_{k=1}^{n} \hat{1}_k^{(t)} \phi_k(u) \Big]^2 \mathrm{d}u = 0 \tag{6.4.10}$$

进而可直接得出

$$\int_0^T \left[1_u^{(t)} - \sum_{k=1}^{n} \hat{1}_k^{(t)}\phi_k(u)\right]^2 du = t - \sum_{k=1}^{n}[\hat{1}_k^{(t)}]^2 \equiv t - \sum_{k=1}^{n}\left(\int_0^t \phi_k(u)du\right)^2 \quad (6.4.11)$$

比较式(6.4.8)与式(6.4.11)，根据式(6.4.10)，有 $\sum_{k=1}^{\infty}\hat{W}_k\int_0^t \phi_k(u)du \to W_t$(m. s.) 成立，即定理证毕。 □

命题 6.4.2 给出了维纳过程在任意完备正交函数集上的表示形式，由此可方便地推导出在加性白噪声中检测高斯信号的似然比。

命题 6.4.3　高斯过程与维纳过程的转换

假设$\{Y_t;\ 0\leqslant t\leqslant T\}$是二阶零均值高斯随机过程，其特征值为$\{\lambda_k\}_{k=1}^{\infty}$且对应的正交特征函数为$\{\psi_k\}_{k=1}^{\infty}$。进一步假设$\{\psi_k\}_{k=1}^{\infty}$完备且$\lambda_k>0$，$k=1, 2, \cdots$，则随机过程为

$$Z_t \triangleq \sum_{k=1}^{\infty}(\hat{Y}_k/\lambda_k^{1/2})\int_0^t \psi_k(u)du, \quad 0\leqslant t\leqslant T \quad (6.4.12)$$

该过程为单位($N_0=2$)维纳过程。其中，$\hat{Y}_k = \int_0^T Y_t\psi_k(t)dt, k = 1,2,\cdots$ 。

证明　显然，命题 6.4.2 所定义的系数序列$\{\hat{W}_k\}_{k=1}^{\infty}$可视为独立同分布于$\mathcal{N}(0, N_0/2)$的随机变量序列。根据 6.2 节中的相关结论，$\{\hat{Y}_k\}_{k=1}^{\infty}$是服从$\mathcal{N}(0, \lambda_k)$分布的独立随机变量序列。因此，$\{\hat{Y}_k/\lambda_k^{1/2}\}_{k=1}^{\infty}$是服从$\mathcal{N}(0, 1)$的独立同分布随机变量。因此，对比式(6.4.12)和式(6.4.7)易知：统计意义上，式(6.4.12)与式(6.4.7)在 $N_0/2=1$ 下一致，即式(6.4.12)中的$\{Z_t;\ 0\leqslant t\leqslant T\}$必是维纳过程。 □

备注　命题 6.4.3 是“白化”后的结果，即它给出了满足约束条件的任意高斯过程转换成维纳过程的方法。实际上，该维纳过程与原过程$\{Y_t;\ 0\leqslant t\leqslant T\}$等价，因为利用 $\hat{Y}_k = \lambda_k^{1/2}\int_0^T \phi_k(t)dZ_t(k = 1,2,\cdots)$ 可得到$\hat{Y}_k$，然后通过 Karhunen-Loéve 展开可得到 Y_t。约束条件$\{\psi_k\}_{k=1}^{\infty}$是完备的，且$\lambda_k>0(k=1, 2, \cdots)$意味着随机过程$\{Y_t;\ 0\leqslant t\leqslant T\}$足以“填满”平方可积函数的空间。又因为$\{Z_t;\ 0\leqslant t\leqslant T\}$是$\{Y_t;\ 0\leqslant t\leqslant T\}$的线性变换，所以在加性维纳过程中推导的结论在满足命题 6.4.3 要求的任意加性高斯噪声下都适用。

6.4.2　白噪声中检测高斯信号

现在考虑用式(6.4.1)所描述的模型检测高斯信号$\{S_t;\ 0\leqslant t\leqslant T\}$的问题。假设随机过程$\{S_t;\ 0\leqslant t\leqslant T\}$均方连续且与$\{W_t;\ 0\leqslant t\leqslant T\}$相互独立，$\{W_t;\ 0\leqslant t\leqslant T\}$的自协方差函数 C_s 的特征值为$\{\lambda_k\}_{k=1}^{\infty}$，相应的正交特征函数为$\{\psi_k\}_{k=1}^{\infty}$。为简化分析，假设$\{\psi_k\}_{k=1}^{\infty}$完备。不过，即使$\{\psi_k\}_{k=1}^{\infty}$不完备，在完备假设下所推导出来的结论也同样适用。

考虑均方和

$$\sum_{k=1}^{\infty}\hat{X}_k\int_0^t \psi_k(u)du \quad (6.4.13)$$

其中，$\hat{X}_k \triangleq \int_0^T \psi_k(t)dX_t, k = 1,2,\cdots$ 。在 H_0 假设下，因为 $X_t=W_t$，所以根据命题 6.4.2 易知序列$\hat{X}_k$ 与 X_t 相同。在 H_1 下有

$$\hat{X}_t = \int_0^T \psi_k(t)d\int_0^t S_u du + \int_0^T \psi_k(t)dW_t$$

$$= \int_0^T \psi_k(t) S_t \mathrm{d}t + \int_0^t \psi_k(t) \mathrm{d}W_t = \hat{S}_k + \hat{W}_k \tag{6.4.14}$$

其中，$\hat{W}_k$ 的定义与命题 6.4.2 相同，$\hat{S}_k$ 是$\{S_t;\ 0\leqslant t\leqslant T\}$在 ψ_k 上的 Karhunen-Loéve 展开系数。由于 Karhunen-Loéve 变换在时间 t 上一致收敛，因此有

$$\sum_{k=1}^{\infty} \hat{S}_k \int_0^t \psi_k(u) \mathrm{d}u = \int_0^t \sum_{k=1}^{\infty} \hat{S}_k \psi_k(u) \mathrm{d}u = \int_0^t S_u \mathrm{d}u \tag{6.4.15}$$

即在 H_1 下，式(6.4.13)可有效表示 X_t。

方程(6.4.13)将连续时间观测值$\{X_t;\ 0\leqslant t\leqslant T\}$化简为等价的离散时间观测序列$\{\hat{X}_k\}_{k=1}^{\infty}$。显然，这个序列在两种假设下均为独立高斯随机变量，差异仅在于 H_0 假设下服从$\hat{X}_k \sim \mathcal{N}(0,\ N_0/2)$分布，$H_1$ 假设下服从$\hat{X}_k \sim \mathcal{N}(0,\ \lambda_k + N_0/2)$分布。因此，式(6.4.1)描述的高斯信号检测问题等价为

$$\begin{aligned} &H_0: \hat{X}_k \sim \mathcal{N}\left(0, \frac{N_0}{2}\right), \qquad k = 1,2,\cdots \\ &H_1: \hat{X}_k \sim \mathcal{N}\left(0, \lambda_k + \frac{N_0}{2}\right), \quad k = 1,2,\cdots \end{aligned} \tag{6.4.16}$$

至此，可以应用 Grenander 定理研究它的似然比和奇异性。

命题 6.4.4　白噪声中高斯信号的似然比

考虑高斯信号$\{S_t;\ 0\leqslant t\leqslant T\}$下式(6.4.1)的假设检验问题。有 $P_0 \equiv P_1$ 且

$$\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(X_0^T) = \left(\prod_{k=1}^{\infty} \frac{N_0/2}{\lambda_k + N_0/2}\right) \times \exp\left(\frac{1}{N_0} \sum_{k=1}^{\infty} \lambda_k \hat{X}_k^2 \Big/ \left(\lambda_k + \frac{N_0}{2}\right)\right) \tag{6.4.17}$$

其中，求和运算指均方和。

证明　利用 Grenander 定理来证明。式(6.4.16)所描述的基于初始 n 个观测值$\hat{X}_1$，…，$\hat{X}_n$ 的检测问题是第 3 章[见式(3.2.85)]所讨论的在加性噪声向量中检测高斯信号向量的特例。因此，基于$\hat{X}_1$，$\hat{X}_2$，…，$\hat{X}_n$ 的对数似然比为

$$\log f_n(X_0^T) = -\sum_{k=1}^{n} \log(1 + 2\lambda_k/N_0) + \frac{1}{N_0} \sum_{k=1}^{n} \lambda_k \hat{X}_k^2 \Big/ \left(\lambda_k + \frac{N_0}{2}\right) \tag{6.4.18}$$

利用不等式 $0\leqslant \log(1+x) \leqslant x(x\geqslant 0)$，对于任意 k 有 $\lambda_k \geqslant 0$ 成立，可得

$$0 \leqslant \sum_{k=1}^{\infty} \log(1 + 2\lambda_k/N_0) \leqslant \frac{2}{N_0} \sum_{k=1}^{\infty} \lambda_k < \infty \tag{6.4.19}$$

因此，只要式(6.4.18)的第二项收敛，就能够证明 $f_n(X_0^T)$依概率收敛。根据柯西准则[见式(6.3.17)]，在 H_1 下，该项在均方意义上收敛的条件是

$$\lim_{n\to\infty} \sup_{m\geqslant n} E_1\left[\left(\sum_{k=n+1}^{m} \lambda_k \hat{X}_k^2 \Big/ \left(\lambda_k + \frac{N_0}{2}\right)\right)^2\right] = 0 \tag{6.4.20}$$

进而可以得到

$$\begin{aligned} &E_1\left[\left(\sum_{k=n+1}^{m} \lambda_k \hat{X}_k^2 \Big/ \left(\lambda_k + \frac{N_0}{2}\right)\right)^2\right] \\ &= \mathrm{var}_1\left(\sum_{k=n+1}^{m} \lambda_k \hat{X}_k^2 \Big/ \left(\lambda_k + \frac{N_0}{2}\right)\right) + \left[E_1\left(\sum_{k=n+1}^{m} \lambda_k \hat{X}_k^2 \Big/ \left(\lambda_k + \frac{N_0}{2}\right)\right)\right]^2 \end{aligned}$$

$$= \sum_{k=n+1}^{m} \lambda_k^2 \mathrm{var}_1(\hat{X}_k^2) / \left(\lambda_k + \frac{N_0}{2}\right)^2 + \left(\sum_{k=n+1}^{m} \lambda_k E\{\hat{X}_k^2\} / \left(\lambda_k + \frac{N_0}{2}\right)\right)^2 \tag{6.4.21}$$

其中，在第二个等式中利用了$\hat{X}_{n+1}$，…，$\hat{X}_m$ 的独立性。在 H_1 假设下，$X_k \sim \mathcal{N}(0, \lambda_k + N_0/2)$，即 $\mathrm{var}_1(\hat{X}_k^2) = 2(\lambda_k + N_0/2)^2$ 和 $E_1\{\hat{X}_k^2\} = \lambda_k + N_0/2$。代入后，式(6.4.20)的左边变为

$$\lim_{n\to\infty} \sup_{m \geqslant n} \left[2 \sum_{k=n+1}^{m} \lambda_k^2 + \left(\sum_{k=n+1}^{m} \lambda_k\right)^2\right] = \lim_{n\to\infty} \left[2 \sum_{k=n+1}^{m} \lambda_k^2 + \left(\sum_{k=n+1}^{m} \lambda_k\right)^2\right] \tag{6.4.22}$$

事实上，在 $\lambda_k \geqslant 0$ 条件下 $\sum_{k=1}^{\infty} \lambda_k < \infty$ 也意味着 $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$，因此式(6.4.22)中的极限为零。

既然 H_1 假设下 $\log f_n(X_0^T)$ 均方收敛，即 $\log f_n(X_0^T)$ 和 $f_n(X_0^T)$ 均依概率收敛，那么根据 Grenander 定理，有 $P_1 \ll P_0$ 成立且 $\mathrm{d}P_1/\mathrm{d}P_0(X_0^T)$ 由式(6.4.17)给出。如果能证明 $P_0 \ll P_1$ 成立，就完成了定理的整个证明过程。事实上，在 Grenander 定理中互换 H_0 和 H_1 的定义后，即可完成 $P_0 \ll P_1$ 的证明。 □

似然比中依赖于观测值的一项为

$$\sum_{k=1}^{\infty} \lambda_k (\hat{X}_k)^2 / \left(\lambda_k + \frac{N_0}{2}\right) \triangleq T_Q \tag{6.4.23}$$

由于 $\hat{X}_k = \int_0^T \psi_k(t) \mathrm{d}X_t$，可以把 T_Q 重新写成

$$T_Q = \int_0^T \int_0^T Q(t,u) \mathrm{d}X_t \mathrm{d}X_u \tag{6.4.24}$$

其中

$$Q(t,u) \triangleq \sum_{k=1}^{\infty} \left(\frac{\lambda_k}{\lambda_k + \frac{N_0}{2}}\right) \psi_k(t) \psi_k(u), \quad 0 \leqslant t, u \leqslant T \tag{6.4.25}$$

进一步假设 $\mathrm{d}X_t = Y_t \mathrm{d}t$ 成立，其中 Y_t 表示信号加白噪声，那么假设检验的统计量 T_Q 变为

$$T_Q = \int_0^T \int_0^T Q(t,u) Y_t Y_u \mathrm{d}t \mathrm{d}u \tag{6.4.26}$$

它是 Y_0^T 的二次型。因此，与向量观测值的例子一样[见式(3.2.86)]，在加性白噪声中检测高斯信号的最优结构仍然是观测值的二次函数。

式(6.4.25)中的函数 Q 是下述积分方程的解

$$C_s(t,u) = \int_0^T Q(t,s) C_s(s,u) \mathrm{d}s + \frac{N_0}{2} Q(t,u), \quad 0 \leqslant t, u \leqslant T \tag{6.4.27}$$

与之对应，离散向量观测模型中二次统计量$\{\boldsymbol{y}^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{y}\}$中的矩阵 $\boldsymbol{Q}$ 满足 $\boldsymbol{\Sigma}_S = \boldsymbol{Q}\boldsymbol{\Sigma}_S + \sigma^2 \boldsymbol{Q}$，其中 σ^2 是式(3.2.85)中白噪声的方差。将矩阵方程改写为

$$(\boldsymbol{\Sigma}_S)_{k,l} = \sum_{j=1}^{m} \boldsymbol{Q}_{k,j} (\boldsymbol{\Sigma}_S)_{j,l} + \sigma^2 \boldsymbol{Q}_{k,l}, \quad 1 \leqslant k, l \leqslant n \tag{6.4.28}$$

易知它对应式(6.4.27)的离散形式。这再次说明在加性白噪声中检测连续波形与检测信号向量类似。

此外，式(6.4.27)是基于 Y_0 估计 S_t 的 Wiener-Hopf 方程(6.3.18)的连续-时间形式。特别地，$\widetilde{s}_t \triangleq \int_0^T Q(t,u) Y_u \mathrm{d}u$[或者更恰当一点，定义为$\int_0^T Q(t,u) \mathrm{d}X_u$]是 H_1 假设下[⊖] S_t 的线性最

⊖ 将在第 7 章讨论该估计问题。

小均方误差估计。因此，T_Q 可视为相关估计器，即

$$T_Q = \int_0^T \widetilde{S}_t Y_t \mathrm{d}t \tag{6.4.29}$$

它通过对观测到的波形和信号的估计值进行相关处理来检测信号。注意，$\widetilde{S}_t$ 不是因果估计，所以式(6.4.29)不是估计器的首选。然而，后续部分将会讨论，似然比表达示(6.4.17)也可表示为[⊖]

$$\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(X_0^T) = \exp\left(\frac{2}{N_0}\int_0^T \hat{S}_t \mathrm{d}X_t - \frac{1}{N_0}\int_0^T (\hat{S}_t)^2 \mathrm{d}t\right) \tag{6.4.30}$$

式中，$\hat{S}_t$ 是 S_t 的因果最小均方误差估计。根据 Cameron-Martin 公式，用随机信号$\{S_t;0\leqslant t\leqslant T\}$的因果估计$\{\hat{S}_t;0\leqslant t\leqslant T\}$替换确知信号$\{S_t;0\leqslant t\leqslant T\}$后，式(6.4.30)就是白噪声中检测确知信号的似然比。

只有在一些特殊情况下才能计算出似然比检测的误差概率，这与离散时间问题类似。因此，在性能分析过程中需要利用误差上界、下界或近似[参见 Mazo 和 Salz(1965)]。二次检测器的性能有时也用所谓的广义信噪比或者偏差来评估，定义为

$$\frac{[E_1\{T_Q(X_0^T)\} - E_0\{T_Q(X_0^T)\}]^2}{\mathrm{var}_0[T_Q(X_0^T)]} \tag{6.4.31}$$

虽然它与误差概率没有直接关系，但是可用来衡量二次型统计量是否能有效区分两种假设[参见 Barton 和 Poor(1992)]。

下面进一步讨论高斯噪声中检测高斯信号的奇异性问题。检测确定信号时，P_0 和 P_1 要么等价，要么奇异。还可发现，在白噪声中检测均方连续信号时，通常 P_0 和 P_1 也是等价的。在这些检测问题中，只有 $P_0 \equiv P_1$ 或 $P_0 \perp P_1$ 两种情况并非偶然。特别是当 P_0 和 P_1 是任意的两个高斯测量时(即在这个测量下观测值服从高斯分布)，总能得到 $P_0 \equiv P_1$ 或者 $P_0 \perp P_1$，而没有其他中间结果。这种分裂性最早由 Feldman(1958)和 Hajek(1958)发现，与之相关的讨论可参考 Grenander(1981)。

利用具有有理谱的观察过程能够很好地揭示高斯噪声中检测高斯信号的奇异性。连续时间随机过程具有有理谱是指其自协方差函数可表示为

$$C(t,s) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}w(t-s)} \phi(w) \mathrm{d}w \tag{6.4.32}$$

其中函数 $\phi(w)$ 具有如下形式

$$\phi(w) = \frac{\sum_{k=1}^{m} a_k w^{2k}}{\sum_{k=1}^{n} b_k w^{2k}}, \quad -\infty < w < \infty \tag{6.4.33}$$

式中，a_k 和 b_k 是常数。与该过程有关的奇异性结论可参考 Slepian(1958)、Feldman(1960)和其他文献。在 H_0 和 H_1 假设下，高斯过程 Y_0^T 的有理谱分别是 ϕ_0 和 ϕ_1，那么，当且仅当下式成立时，$P_0 \equiv P_1$ 成立

⊖ 证明式(6.4.30)的有效性需要重新理解已定义的积分 $\int_0^T \hat{S}_t \mathrm{d}X_t$，将在 6.4.3 节中讨论。

$$\lim_{w\to\infty}[\phi_1(w)/\phi_0(w)] = 1 \tag{6.4.34}$$

举例来说，假设有观测模型

$$\begin{aligned} H_0: Y_t &= N_t, & 0 \leqslant t \leqslant T \\ H_1: Y_t &= S_t + N_t, & 0 \leqslant t \leqslant T \end{aligned} \tag{6.4.35}$$

其中，$\{S_t;\ 0\leqslant t\leqslant T\}$和$\{N_t;\ 0\leqslant t\leqslant T\}$是有理谱分别为 ϕ_S 和 ϕ_N 的独立高斯信号。此时，有 $\phi_0=\phi_N$ 和 $\phi_1=\phi_N+\phi_S$，P_0 和 P_1 等价的充要条件为

$$\lim_{w\to\infty}\phi_S(w)/\phi_N(w) = 0 \tag{6.4.36}$$

对式(6.4.36)的直观解释是：式(6.4.35)所描述的假设检验问题非奇异要求信号比噪声变化慢。换言之，信号变化比噪声慢对应非奇异检测，信号变化比噪声快($\phi_S(w)/\phi_N(w)\to\infty$)导致奇异。不过，在信号和噪声的谱形状相同时($0<\lim_{w\to\infty}\phi_S(w)/\phi_N(w)<\infty$)，利用有限时间区间的观测数据，仍能可靠检测出信号，即 $P_0\perp P_1$。

Shepp(1966)给出了高斯二分法的另一个例子，并讨论了维纳测量的等价性和奇异性。就信号检测而言，假设检验问题为

$$\begin{aligned} H_0: X_t &= W_t, & 0 \leqslant t \leqslant T \\ H_1: X_t &= \int_0^t S_u \mathrm{d}u + W_t, & 0 \leqslant t \leqslant T \end{aligned}$$

式中，$\{W_t;\ 0\leqslant t\leqslant T\}$是维纳过程，$\{S_t;\ 0\leqslant t\leqslant T\}$是均值为$E\{S_t\}$、协方差为$C_S(t,\ u)$的高斯信号。根据 Shepp 的结论：当且仅当$\int_0^T (E\{S_t\})^2\mathrm{d}t<\infty$ 和 $\int_0^T\int_0^T C_S^2(t,u)\mathrm{d}t\mathrm{d}u<\infty$ 时 $P_0\equiv P_1$ 成立；否则，$P_0\perp P_1$。当 $P_0\equiv P_1$ 时，似然比为

$$\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(X_0^T) = \left[\prod_{k=1}^{\infty}\left(\frac{1+2\lambda_k}{N_0}\right)^{-1}\right]\exp\left\{\frac{2}{N_0}g(X_0^T)\right\}$$

其中

$$g(X_0^T) \triangleq \int_0^T E\{S_t\}\mathrm{d}X_t - \frac{1}{2}\int_0^T [E\{S_t\}]^2\mathrm{d}t + \frac{1}{2}\int_0^T\int_0^T Q(t,u)\mathrm{d}X_t\mathrm{d}X_u$$

Q 是式(6.4.27)的解。显然，上式在 $C_S\equiv 0$ 时简化为 Cameron-Martin 公式，在 $E\{S_t\}\equiv 0$ 时简化为命题 6.4.4 的公式。一旦 C_S 连续，则 $\int_0^T\int_0^T C_S^2(t,u)\mathrm{d}t\mathrm{d}u$ 有限，且 $E\{S_t\}$的变差有界时有 $\int_0^T E^2\{S_t\}\mathrm{d}t<\infty$ 成立，Shepp 所给出的结论是之前所讨论的高斯白噪声中检测信号的推广。

6.4.3　随机信号似然比检测的估计-相关器表示

在上一节中，针对假设检验问题

$$\begin{aligned} H_0: X_t &= W_t, & 0 \leqslant t \leqslant T \\ H_1: X_t &= W_t + \int_0^t S_u \mathrm{d}u, & 0 \leqslant t \leqslant T \end{aligned} \tag{6.4.37}$$

其中$\{W_t;\ 0\leqslant t\leqslant T\}$是维纳噪声过程，$\{S_t;\ 0\leqslant t\leqslant T\}$是均方连续的零均值高斯信号。根据命题 6.4.4，似然比为

$$\log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(X_0^T) = \sum_{k=1}^{\infty} \log \frac{N_0/2}{\lambda_k + N_0/2} + \frac{1}{N_0} \sum_{k=1}^{\infty} \lambda_k \hat{X}_k^2/(\lambda_k + N_0/2) \tag{6.4.38}$$

式中，$\hat{X}_k \triangleq \int_0^T \psi_k(t)\mathrm{d}X_t$，$\{\lambda_k\}$和$\{\psi_k\}$是信号自协方差函数的特征值和特征函数。对于式(6.4.37)而言[见式(6.4.30)]，$\mathrm{d}P_1/\mathrm{d}P_0$ 同样能表示成

$$\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(X_0^T) = \frac{2}{N_0}\int_0^T \hat{S}_t \mathrm{d}X_t - \frac{1}{N_0}\int_0^T (\hat{S}_t)^2 \mathrm{d}t \tag{6.4.39}$$

式中，$\hat{S}_t$ 是 H_1 假设下 S_t 的最小均方误差估计。接下来证明式(6.4.39)成立并将其推广到非高斯信号情况。为简化起见，令 $T=N_0/2=1$。

在推导式(6.4.39)之前，有必要深入理解这个公式。对每个时间 t 而言，估计值$\{\hat{S}_t;0\leqslant t\leqslant 1\}$是观测值 X_0^t 的函数，也是一个随机过程。虽然已经定义了形如$\int_0^1 h_t \mathrm{d}X_t$ 的积分形式，但是要求被积函数 h_t 是确知的。因此，为了推出式(6.4.39)，有必要给出被积函数是随机过程时的积分定义，即$\int_0^1 \hat{S}_t \mathrm{d}X_t$ 的具体含意。

实际上，有很多种方法可用来定义$\int_0^1 \hat{S}_t \mathrm{d}X_t$，但是它们并不等价。为了说明定义一个随机过程积分方法的难度，考虑定义$\int_0^1 W_t \mathrm{d}W_t$ 的问题，其中$\{W_t;0\leqslant t\leqslant 1\}$是单位维纳过程。假如采用常规的定义，即

$$\int_0^1 W_t \mathrm{d}W_t \overset{(\mathrm{m.s.})}{=\!=\!=} \lim_{\Delta_n\to 0}\sum_{i=1}^{n} W_{\xi_i^{(n)}}(W_{t_i^{(n)}} - W_{t_{i-1}^{(n)}}) \tag{6.4.40}$$

其中，$0=t_0^{(n)}<t_1^{(n)}<\cdots<t_n^{(n)}=1$，$\xi_i^{(n)}\in[t_{i-1}^{(n)},\ t_i^{(n)}]$，$i=1,\ \cdots,\ n$ 且 $\Delta_n=\max\limits_{1\leqslant i\leqslant n}|t_i^{(n)}-t_{i-1}^{(n)}|$。如果该积分存在，即对于任意选择的 t_i 和 ξ_i，极限都一样，即

$$E\left\{\int_0^1 W_t \mathrm{d}W_t\right\} = \lim_{\Delta_n\to 0}\sum_{i=1}^{n} E\{W_{\xi_i^{(n)}}(W_{t_i^{(n)}} - W_{t_{i-1}^{(n)}})\} = \lim_{\Delta_n\to 0}\sum_{i=1}^{n}(\xi_i^{(n)} - t_{i-1}^{(n)}) \tag{6.4.41}$$

显然，式(6.4.41)最右边的量可在 0～1 之间任意取值，这取决于在区间$[t_{i-1}^{(n)},\ t_i^{(n)}]$上如何选择 $\xi_i^{(n)}$。例如，对于 $i=1,\ \cdots,\ n$，如果选取 $\xi_i^{(n)}=t_{i-1}^{(n)}$，则可以得出 $E\left[\int_0^1 W_t \mathrm{d}W_t\right]=0$；如果选取 $\xi_i^{(n)}=t_i^{(n)}$，则可以得出 $E\left[\int_0^1 W_t \mathrm{d}W_t\right]=1$。因此，式(6.4.40)不具有所期望的性质。

定义随机积分的最核心问题是被积函数和积分函数在统计上可能不独立。就两者都是相同的维纳过程而言，$W_{\xi_i^{(n)}}$ 与增量$(W_{t_i^{(n)}}-W_{t_{i-1}^{(n)}})$的依赖程度与 $\xi_i^{(n)}$ 在间隔$[t_{i-1}^{(n)},\ t_i^{(n)}]$中的位置密切相关。因此，相比确定函数的积分而言，两个随机过程的积分需要更加明确的定义。特别地，应该明确 $\xi_i^{(n)}$ 在$[t_{i-1}^{(n)},\ t_i^{(n)}]$中的具体位置。虽然在选择 $\xi_i^{(n)}$ 时或多或少会根据约定进行，且每种方法都存在这样或那样的不足；但是，最常用的方式是选取 $\xi_i^{(n)}=t_{i-1}^{(n)}$。它在描述信号检测和估计公式时特别有用，这也正是采取这种方法的原因。为此，随机过程$\{Y_t;0\leqslant t\leqslant T\}$关于随机过程$\{X_t;0\leqslant t\leqslant T\}$的积分定义为

$$\int_0^T Y_t \mathrm{d}X_t \overset{(\mathrm{m.s.})}{=} \lim_{\Delta_n\to\infty}\sum_{i=1}^{n} Y_{t_{i-1}^{(n)}}(X_{t_i^{(n)}} - X_{t_{i-1}^{(n)}}) \tag{6.4.42}$$

其中 $t_i^{(n)}$ 和 Δ_n 与前面一样。对任意划分，式(6.4.42)以概率 1 有唯一极限时，积分 $\int_0^1 Y_t \mathrm{d}X_t$ 才存在。

式(6.4.42)中的积分通常称为 Ito 随机积分，推广其定义后，可扩展可积随机过程的类型[参见 Wong 和 hajek(1985)]。不过，式(6.4.42)的定义对所涉及的应用而言已经足够。事实上，根据定义，增量 $\mathrm{d}X_t$ 对应"前向增量"，即 $\mathrm{d}X_t = X_{t+\mathrm{d}t} - X_t$ ⊖。

式(6.4.42)所定义的随机积分具有普通积分的线性性质，即对标量 α 和 β 有 $\int_0^T [\alpha Y_t + \beta Z_t]\mathrm{d}X_t = \alpha\int_0^T Y_t \mathrm{d}X_t + \beta\int_0^T Z_t \mathrm{d}X_t$ 。然而，Ito 随机积分不满足普通微积分的所有规则。

命题 6.4.5　Ito 修正项

假设$\{W_t;\ 0\leqslant t\leqslant 1\}$是单位维纳过程，定义随机过程$\{X_t;\ 0\leqslant t\leqslant 1\}$为

$$X_t = X_0 + \int_0^t \phi_u \mathrm{d}W_u,\quad 0\leqslant t\leqslant 1 \tag{6.4.43}$$

其中 ϕ 是在[0，1]区间上的平方可积函数且 X_0 是随机变量。积分 $\int_0^1 X_t \mathrm{d}X_t$ 存在且可表示为

$$\int_0^1 X_t \mathrm{d}X_t = \frac{1}{2}(X_1^2 - X_0^2) - \frac{1}{2}\int_0^1 \phi_u^2 \mathrm{d}u \tag{6.4.44}$$

证明　对于[0，1]上的一个划分 $t_0^{(n)}$，…，$t_n^{(n)}$，利用 $a(b-a)=(b^2-a^2)/2-(b-a)^2/2$ 后，可得出

$$\begin{aligned}\sum_{i=0}^{n-1} X_{t_i^{(n)}}(X_{t_{i+1}^{(n)}} - X_{t_i^{(n)}}) &= \frac{1}{2}\sum_{i=0}^{n-1} X^2_{t_{i+1}^{(n)}} - \frac{1}{2}\sum_{i=0}^{n-1} X^2_{t_i^{(n)}} - \frac{1}{2}\sum_{i=0}^{n-1}(X_{t_{i+1}^{(n)}} - X_{t_i^{(n)}})^2 \\ &= \frac{1}{2}(X_1^2 - X_0^2) - \frac{1}{2}\sum_{i=0}^{n-1}\left(\int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \phi_u \mathrm{d}W_u\right)^2\end{aligned} \tag{6.4.45}$$

$Y_i = \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \phi_u \mathrm{d}W_u$ ，$i=0$，1，…，$n-1$ 是服从 $\mathcal{N}\left(0,\int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \phi_u^2 \mathrm{d}u\right)$ 分布的独立随机变量。因此，对随机变量序列

$$Z_n = \sum_{i=0}^{n-1}\left(\int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \phi_u \mathrm{d}W_u\right)^2,\quad n = 1,2,\cdots \tag{6.4.46}$$

可得 $E\{Z_n\} = \sum_{i=0}^{n-1}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \phi_u^2 \mathrm{d}u = \int_0^1 \phi_u^2 \mathrm{d}u$ ，并且

$$E\left\{\left(Z_n - \int_0^1 \phi_u^2 \mathrm{d}u\right)^2\right\} = \mathrm{var}(Z_n) = \sum_{i=0}^{n-1}\mathrm{var}\left[\left(\int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \phi_u \mathrm{d}W_u\right)^2\right] \tag{6.4.47}$$

如果 $X\sim\mathcal{N}(0,\ \sigma^2)$，则 $\mathrm{var}(X^2)=2\sigma^4$。因此

$$E\left\{\left(Z_n - \int_0^1 \phi_u^2 \mathrm{d}u\right)^2\right\} = 2\sum_{i=0}^{n-1}\left(\int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \phi_u \mathrm{d}u\right)^2 \leqslant 2\delta_n \sum_{i=0}^{n-1}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \phi_u^2 \mathrm{d}u = 2\delta_n\int_0^1 \phi_u^2 \mathrm{d}u$$

其中

⊖ 随机积分的另一种定义是 Stratonovich 随机积分，它把 $\mathrm{d}X_t$ 解释为后向增量($\xi_i^{(n)} = t_i^{(n)}$)。这个积分性质与 Ito 积分稍有不同。两种定义各有优缺点。

$$\delta_n \triangleq \max_{0\leqslant i\leqslant n-1}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \phi_u^2 \mathrm{d}u \tag{6.4.48}$$

结合式(6.4.45)～式(6.4.48)，得到

$$\sum_{i=0}^{n-1} X_{t_i^{(n)}}(X_{t_{i+1}^{(n)}} - X_{t_i^{(n)}}) = \frac{1}{2}(X_1^2 - X_0^2) - \frac{1}{2}\int_0^1 \phi_u^2 \mathrm{d}u + \varepsilon_n$$

其中 $E(\varepsilon_n^2)\leqslant 2\delta_n$。$\Delta_n \to 0$ 意味着 $\delta_n \to 0$。这样，当 $\Delta_n \to 0$ 时，$\sum_{i=0}^{n-1} X_{t_i^{(n)}}(X_{t_{i+1}^{(n)}} - X_{t_i^{(n)}})$ 均方收敛，并且它的极限由式(6.4.44)给出。 □

需要注意的是：式(6.4.44)和从普通微积分得到的相关表达式存在差异。特别地，如果 f 是[0，1]上有界变分的连续函数，可得到更一般的表达式，即

$$\int_0^1 f_t \mathrm{d}f_t = \frac{1}{2}(f_1^2 - f_0^2) \tag{6.4.49}$$

式(6.4.44)中的“修正项”涉及了有关$\{X_t;\ 0\leqslant t\leqslant 1\}$的二阶变分，即

$$\int_0^1 (\mathrm{d}X_t)^2 = \lim_{\Delta_n \to 0}\sum_{i=0}^{n-1}(X_{t_{i+1}^{(n)}} - X_{t_i^{(n)}})^2 = \int_0^1 \phi_t^2 \mathrm{d}t \tag{6.4.50}$$

对于一个有界变分的连续函数，因为$(\mathrm{d}f_t)^2=(\mathrm{d}t)^2$，所以二阶变分为零。然而，对于形如式(6.4.43)的随机过程，$(\mathrm{d}X_t)^2$ 的特性与 $\phi_t^2\mathrm{d}t$ 一样。在第 7 章中还会进一步讨论该问题。同时，命题 6.4.5 包含了所需要证明的随机积分的性质。

现在继续讨论式(6.4.38)中的似然比。注意，在命题 6.2.1 中，Radon-Nikodym 导数 $\mathrm{d}P_1/\mathrm{d}P_0$ 利用其对 P_0 的积分来定义。因此，要证明式(6.4.39)就是似然比的表达式，就仅需证明式(6.4.38)和式(6.4.39)以概率 1 等于 P_0。这也意味着它们以概率 1 等于 P_1，因为 $P_1 \ll P_0$。

变量$\hat{X}_k$ 由 $\hat{X}_k = \int_0^1 \psi_k(t)\mathrm{d}X_t$ 给定，而且在 P_0 下$\{X_t;\ 0\leqslant t\leqslant 1\}$是维纳过程。应用命题 6.4.5后，有

$$\begin{aligned}\hat{X}_k &= 2\int_0^1\left[\int_0^t \psi_k(u)\mathrm{d}X_u\right]\mathrm{d}\left[\int_0^t \psi_k(u)\mathrm{d}X_u\right] + \int_0^1 \psi_k^2(t)\mathrm{d}t \\ &= 2\int_0^1\left[\int_0^t \psi_k(u)\mathrm{d}X_u\right]\psi_k(t)\mathrm{d}X_t + \int_0^1 \psi_k^2(t)\mathrm{d}t\end{aligned} \tag{6.4.51}$$

因为 $\mathrm{d}\int_0^t \psi_k(u)\mathrm{d}W_u = \psi_k(t)\mathrm{d}W_t$，所以 P_0 的概率为 1。根据 Mercer 定理和 $0 \leqslant \lambda_k/(\lambda_k+1)\leqslant \lambda_k$，有 $\sum_{k=1}^{\infty}[\lambda_k/(\lambda_k+1)]\psi_k(t)\psi_t(u)$ 一致收敛。结合式(6.4.51)后，有

$$\log\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(X_0^1) = \sum_{k=1}^{\infty}\log[1/(1+\lambda_k)] + \int_0^1\left[\int_0^t Q(t,u)\mathrm{d}X_u\right]\mathrm{d}X_t + \frac{1}{2}\int_0^1 Q(t,t)\mathrm{d}t \tag{6.4.52}$$

与式(6.4.22)类似，Q 定义为

$$Q(t,u) = \sum_{k=1}^{\infty}\frac{\lambda_k}{1+\lambda_k}\psi_k(t)\psi_k(u),\quad 0\leqslant t,u\leqslant 1$$

在 6.3 节中，Q 满足积分方程

$$C_s(t,u) = \int_0^1 Q(t,s)C_s(s,u)\mathrm{d}s + Q(t,u),\quad 0\leqslant t,u\leqslant 1 \tag{6.4.53}$$

在第 7 章中，将证明该式是根据 $\widetilde{S}_t=\int_0^1 Q(s,u)\mathrm{d}X_u$ 和 $\int_0^u S_v\mathrm{d}v+W_u, 0\leqslant u\leqslant 1$ 估计 S_t 的 Wiener-Hopf 方程。函数 Q 通常可视为 C_s 的 Fredholm 分解。一般而言，$\widetilde{S}_t$ 是信号的非因果估计。从 $\int_0^u S_v\mathrm{d}v+W_u$ ，$0\leqslant u\leqslant 1$ 中得到的 S_t 的最小均方误差因果估计量为[⊖]

$$\hat{S}_t=\int_0^t h(t,u)\mathrm{d}X_u,\quad 0\leqslant t\leqslant 1 \tag{6.4.54}$$

其中 h 是因果 Wiener-Hopf 方程的解，即

$$C_s(t,u)=\int_0^t h(t,s)C_s(s,u)\mathrm{d}s+h(t,u),\quad 0\leqslant u\leqslant t\leqslant 1 \tag{6.4.55}$$

可以证明，上述方程有唯一连续解。与第 5 章中的 Wiener-Kolmogorov 方程类似，虽然仅对 Q 进行简单的截断操作不能得到 h，但是，按照下式可从 h 中获得 Q(具体推导留作练习)

$$Q(t,u)=h(t,u)+h(u,t)-\int_0^1 h(s,t)h(s,u)\mathrm{d}s,\quad 0\leqslant t,u\leqslant 1 \tag{6.4.56}$$

其中，若 $y\geqslant x$，则 $h(x,\ y)=0$。

将式(6.4.56)代入式(6.4.52)，可得

$$\log\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(X_0^1)=\sum_{k=1}^{\infty}\log[1/(1+\lambda_k)]+\int_0^1\int_0^t h(t,u)\mathrm{d}X_u\mathrm{d}X_t-\int_0^1\left[\int_0^t\left\{\int_0^1 h(s,t)h(s,u)\mathrm{d}s\right\}\mathrm{d}X_u\right]\mathrm{d}X_t$$
$$+\int_0^1 h(t,t)\mathrm{d}t-\frac{1}{2}\int_0^1\int_0^1 h^2(s,t)\mathrm{d}s\mathrm{d}t \tag{6.4.57}$$

考虑式(6.4.57)右边的中间项，交换积分顺序，并利用 $t>s$ 时 $h(s,\ t)=0$，根据命题 6.4.5，有

$$\int_0^1\left[\int_0^s\left\{\int_0^t h(s,u)\mathrm{d}X_u\right\}h(s,t)\mathrm{d}X_t\right]\mathrm{d}s=\frac{1}{2}\int_0^1\left[\int_0^s h(s,u)\mathrm{d}X_u\right]^2\mathrm{d}s-\frac{1}{2}\int_0^1\int_0^s h^2(s,u)\mathrm{d}u\mathrm{d}s \tag{6.4.58}$$

结合式(6.4.54)、式(6.4.57)和式(6.4.58)，可得到

$$\log\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(X_0^1)=\int_0^1\hat{S}_t\mathrm{d}X_t-\frac{1}{2}\int_0^1[\hat{S}_t]^2\mathrm{d}t+\left[\int_0^1 h(t,t)-\sum_{k=1}^{\infty}\log(1+\lambda_k)\right] \tag{6.4.59}$$

$d(\lambda)\triangleq\prod_{k=1}^{\infty}(\lambda^{-1}\lambda_k+1)$ 也称为 C_s 的 Fredholm 行列式，而且可以表示为[参见 Kailath (1969)]

$$\log d(\lambda)=\int_0^1 h_\lambda(t,t)\mathrm{d}t \tag{6.4.60}$$

其中 h_λ 是下述方程的解

$$C_s(t,u)=\int_0^1 h_\lambda(t,s)C_s(s,u)\mathrm{d}s+\lambda^{-1}h_\lambda(t,u),\quad 0\leqslant t\leqslant 1 \tag{6.4.61}$$

将 $\lambda=1$ 代入式(6.4.61)，可得到 $\prod_{k=1}^{\infty}(\lambda^{-1}\lambda_k+1)=\int_0^1 h(t,t)\mathrm{d}t$ ，并且式(6.4.59)的似然比为

⊖ 将在第 7 章中详细推导。

$$\log \frac{\mathrm{d}P_1}{\mathrm{d}P_0}(X_0^1) = \int_0^1 \hat{S}_t \mathrm{d}X_t - \frac{1}{2}\int_0^1 [\hat{S}_t]^2 \mathrm{d}t \tag{6.4.62}$$

式(6.4.62)的似然比很有趣，用估计信号$\{\hat{S}_t;\ 0\leqslant t\leqslant 1\}$代替已知信号后，在白噪声中检测高斯信号$\{S_t;\ 0\leqslant t\leqslant 1\}$的似然比和在白噪声中检测已知信号的似然比完全一样。上述推导过程在解释原因方面还不够直观，下面结合高斯信号的新息定理来进一步说明。

命题 6.4.6　高斯过程的新息定理

在 H_1 假设下，观测模型为

$$X_t = \int_0^T S_u \mathrm{d}u + W_t,\quad 0\leqslant t\leqslant 1 \tag{6.4.63}$$

其中$\{S_u;\ 0\leqslant u\leqslant 1\}$是独立于$\{W_t;\ 0\leqslant t\leqslant 1\}$的均方连续高斯信号。可以将观测模型等价为

$$X_t = \int_0^t \hat{S}_u \mathrm{d}u + I_t,\quad 0\leqslant t\leqslant 1 \tag{6.4.64}$$

其中，$\{I_t;\ 0\leqslant t\leqslant 1\}$是单位维纳过程，且对所有的 $1\geqslant t\geqslant s\geqslant 0$，$I_t - I_s$ 独立于$\{X_u;\ 0\leqslant u\leqslant s\}$。

证明　根据定义，显然有 $I_t = W_t + \int_0^t (S_u - \hat{S}_u)\mathrm{d}u$ 。因为$\{S_t;\ 0\leqslant t\leqslant 1\}$和$\{W_t;\ 0\leqslant t\leqslant 1\}$都是零均值高斯过程，且零均值高斯过程$\{\hat{S}_t;\ 0\leqslant t\leqslant 1\}$是 $S_0^1 + W_0^1$ 的线性变换，所以$\{I_t;\ 0\leqslant t\leqslant 1\}$也是零均值高斯过程。为了证明$\{I_t;\ 0\leqslant t\leqslant 1\}$是维纳过程，仅需要证明 $E\{I_t I_s\} = \min\{t,\ s\}$成立。

考虑到 $E\{(I_t - I_s)^2\}$对于所有 $1\geqslant t\geqslant s\geqslant 0$ 成立，可得

$$\begin{aligned} E\{(I_t - I_s)^2\} = {} & E\{(W_t - W_s)^2\} + 2\int_s^t E\{(W_t - W_s)(S_u - \hat{S}_u)\}\mathrm{d}u \\ & + \int_s^t\int_s^t E\{(S_u - \hat{S}_u)(S_v - \hat{S}_v)\}\mathrm{d}u\mathrm{d}v \end{aligned} \tag{6.4.65}$$

因为$\{W_t;\ 0\leqslant t\leqslant 1\}$是维纳过程，所以式(6.4.65)右边的第一项等于$(t-s)$。考虑式(6.4.65)第二项积分中的期望，可表示为

$$\begin{aligned} E\{(W_t - W_s)(S_u - \hat{S}_u)\} &= E\{(W_t - W_u)(S_u - \hat{S}_u)\} + E\{(W_u - W_s)(S_u - \hat{S}_u)\} \\ &= E\{(W_t - W_u)\}E\{(S_u - \hat{S}_u)\} + E\{(W_u - W_s)(S_u - \hat{S}_u)\} \\ &= E\{(W_u - W_s)(S_u - \hat{S}_u)\} \end{aligned} \tag{6.4.66}$$

第二个等式成立的原因是$\hat{S}_u$ 只与 S_0^u 和 W_0^u 有关，而 S_0^u 和 W_0^u 都与$(W_t - W_u)$相互独立。第三个等式成立是因为 $E\{W_t - W_s\} = E\{S_u - \hat{S}_u\} = 0$。注意，式(6.4.65)右边的第三项中的被积函数关于$(u,\ v)$对称，综上所述，可写成

$$\begin{aligned} E\{(I_t - I_s)^2\} &= (t-s) + 2\int_0^t E\left\{(S_u - \hat{S}_u)\left(W_u - W_s + \int_s^u (S_v - \hat{S}_v)\mathrm{d}v\right)\right\}\mathrm{d}u \\ &= (t-s) + 2\int_0^t E\{(S_u - \hat{S}_u)(I_u - I_s)\}\mathrm{d}u \end{aligned} \tag{6.4.67}$$

注意，对于所有的 $u\geqslant s$，$I_u - I_s$ 是 X_0^u 的线性函数，因此，对于所有 $u\geqslant s$，根据正交准则有

$$E\{(S_u - \hat{S}_u)(I_u - I_s)\} = 0$$

进而对所有的 $1\geqslant t\geqslant s\geqslant 0$ 都有 $E\{(I_t-I_s)^2\}=t-s$ 成立。

又因为 $(I_t-I_s)=\frac{1}{2}\{(I_t-I_0)^2+(I_s-I_0)^2-(I_t-I_s)^2\}$，有

$$E\{I_tI_s\}=\frac{1}{2}[E\{(I_t-I_0)^2\}+E\{(I_s-I_0)^2\}-E\{(I_t-I_s)^2\}$$

$$=\frac{1}{2}[t+s-|t-s|]=\min\{t,s\},\quad 0\leqslant t,s\leqslant 1$$

因此，$\{I_t;\ 0\leqslant t\leqslant 1\}$ 是维纳过程。

为了验证对所有的 $1\geqslant t\geqslant s\geqslant 0$，都有 (I_t-I_s) 独立于 X_0^s，需要注意，对 $1\geqslant t\geqslant s\geqslant 0$，有

$$E\{(I_t-I_s)X_u\}=E\{(W_t-W_s)X_u\}+\int_s^t E\{(S_v-\hat{S}_v)X_u\}\mathrm{d}v=0$$

这里利用了 W_t-W_s 独立于 X_u 并且对于所的 $s\leqslant v\leqslant t$ 都有 $S_v-\hat{S}_v$ 与 X_u 正交这一事实。因此，对于所有的 $u\leqslant s$，I_t-I_s 和 X_u 正交。又因为它们都是高斯过程，所以 (I_t-I_s) 和 X_0^s 相互独立。 □

命题 6.4.6 中定义的过程 $\{I_t;\ 0\leqslant t\leqslant 1\}$ 通常称为 $\{X_t;\ 0\leqslant t\leqslant 1\}$ 的新息过程。在每个时刻，和在第 5 章中讨论的离散-时间相似，它表示观测值中不能由过去预测的部分。为了更好地说明新息过程的性质，将观测值在 $(t,\ t+\mathrm{d}t)$ 内的变化表示成

$$X_{t+\mathrm{d}t}-X_t=\int_t^{t+\mathrm{d}t}\hat{S}_u\mathrm{d}u+(I_{t+\mathrm{d}t}+I_t)$$

由于 $h(t,\ u)$ 连续，过程 $\{\hat{S}_t;\ 0\leqslant t\leqslant 1\}$ 的采样路径也连续。在 $\mathrm{d}t\to 0$ 时，可得出

$$X_{t+\mathrm{d}t}-X_t\sim\hat{S}_t\mathrm{d}t+(I_{t+\mathrm{d}t}+I_t)\tag{6.4.68}$$

显然，观测值在 $(t,\ t+\mathrm{d}t)$ 的变化由依赖过去 X_0^t 的可预测 $\hat{S}_t\mathrm{d}t$ 分量和完全独立于过去的新息分量 $I_{t+\mathrm{d}t}-I_t$ 组成。

利用新息后，可更为简洁地推导出式(6.4.39)中的估计器-相关器结构。特别地，假设对所有的 $0\leqslant t\leqslant 1$，利用 t 之前的观察数据得到的似然比为 $L_t(X_0^1)$，即 $L_t(X_0^1)$ 只与 X_0^t 相关。对 $1/\mathrm{d}t$ 是整数的每个 $\mathrm{d}t>0$ 而言，可将似然比写成

$$\log L_t(X_0^1)=\sum_{k=1}^{1/\mathrm{d}t}[\log L_{k\mathrm{d}t}(X_0^1)-\log L_{(k-1)\mathrm{d}t}(X_0^1)]\tag{6.4.69}$$

这里，已经从 $\log L_t(X_0^1)$ 中加上和减去了完全相同的项。现在，对于无穷小的 $\mathrm{d}t$，过去的观测值 X_0^t 和当前观测值 $X_0^{t+\mathrm{d}t}$ 可视为对随机过程 $\{Y_u;\ t\leqslant u\leqslant t+\mathrm{d}t\}$ 进行取样。根据式(6.4.68)，该随机过程可表示为

$$\begin{aligned}H_0:Y_u&=W_{t+u}-W_t, & t\leqslant u\leqslant t+\mathrm{d}t\\H_1:Y_u&=\hat{S}_t\mathrm{d}u+(I_{t+u}-I_t), & t\leqslant u\leqslant t+\mathrm{d}t\end{aligned}\tag{6.4.70}$$

一方面，在给定 X_0^t 后，$\hat{S}_t$ 已知；另一方面，在 $t\leqslant u\leqslant t+\mathrm{d}t$ 上，$W_{t+u}-W_t$ 和 $I_{t+u}-I_t$ 是相同的维纳过程。因此，在获得当前观测 $X_0^{t+\mathrm{d}t}$ 后，似然比的增量可由已知白噪声中信号的 Cameron-Martin 公式给出，即

$$\log L_{t+\mathrm{d}t}(X_0^1)-\log L_t(X_0^1)\sim\int_t^{t+\mathrm{d}t}\hat{S}_t\mathrm{d}X_u-\frac{1}{2}\int_t^{t+\mathrm{d}t}(\hat{S}_t)^2\mathrm{d}u=\hat{S}_t(X_{t+\mathrm{d}t}-X_t)-\frac{1}{2}(\hat{S}_t)^2\mathrm{d}t\tag{6.4.71}$$

将上式代入式(6.4.69)，可以得到

$$\log L_1(X_0^1) \sim \sum_{k=1}^{1/\mathrm{d}t}[\hat{S}_{(k-1)\mathrm{d}t}(X_{k\mathrm{d}t}-X_{(k-1)\mathrm{d}t})-\frac{1}{2}\hat{S}_{(k-1)\mathrm{d}t}^2\mathrm{d}t] \xrightarrow{\mathrm{d}t\to 0} \int_0^1 \hat{S}_t\mathrm{d}X_t-\frac{1}{2}\int_0^1(\hat{S}_t)^2\mathrm{d}t \tag{6.4.72}$$

以上推导过程可用于解释估计-相关器的似然比计算式的有效性，且清晰地展示了在似然比计算式中出现了 Ito 积分项的原因。特别地，因为 $\hat{S}_t$ 连续，也可将近似表达式(6.4.68)替代为

$$X_{t+\mathrm{d}t}-X_t \sim \hat{S}_{t_0}\mathrm{d}t+(I_{t+\mathrm{d}t}-I_t) \tag{6.4.73}$$

其中 t_0 是位于 t 与$(t+\mathrm{d}t)$之间的任意数。不过，仅当选择 $t_0=t$ 时，$\hat{S}_{t0}$才完全依赖于 X_0^t，且式(6.4.71)才有效。选择 $t_0=t$ 后，$\int_0^1 \hat{S}_t\mathrm{d}X_t$ 可理解为 $\int_0^1 \hat{S}_t(X_{t+\mathrm{d}t}-X_t)$，即 Ito 积分。

下面给出估计-相关器的似然比表达式的更为严格的证明过程。虽然新息过程 $I_t \triangleq X_t-\int_0^1 \hat{S}_u\mathrm{d}u$，$0\leqslant t\leqslant 1$ 具有维纳统计特性是采用估计-相关器结构的根本原因，但是在不考虑$\hat{S}_u$ 具有连续性和$\{I_t-I_s;\ s\leqslant t\leqslant 1\}$是独立于 X_0^s 的维纳过程时，估计-相关器的推导过程也同样有效。事实上，新息的性质不仅适用于高斯信号模型，在比高斯模型更广泛的信号模型中也是有效的。

特别地，考虑到假设检验问题

$$\begin{aligned} H_0&:X_t=W_t, && 0\leqslant t\leqslant 1\\ H_1&:X_t=\int_0^t S_u\mathrm{d}u+W_t, && 0\leqslant t\leqslant 1 \end{aligned} \tag{6.4.74}$$

这里，$\{S_t;\ 0\leqslant t\leqslant 1\}$表示信号过程(未必是高斯过程)，满足条件

$$E\left\{\int_0^1 |S_t|\mathrm{d}t<\infty\right\} \tag{6.4.75a}$$

和

$$\int_0^1 S_t^2\mathrm{d}t<\infty,\quad \text{w. p. 1} \tag{6.4.75b}$$

$\{W_t;\ 0\leqslant t\leqslant 1\}$是单位维纳过程，对每个 $s\in[0,\ 1]$，过程$\{W_t-W_s;\ s\leqslant t\leqslant 1\}$和$\{X_0^s,\ W_0^s\}$相互独立。如果信号独立于噪声，后一个条件可以自动满足。在这些假设下，新息可表示为

$$I_t \triangleq X_t-\int_0^t \hat{S}_u\mathrm{d}u \tag{6.4.76}$$

其中$\hat{S}_t\triangleq E_1\{S_t|X_0^t\}$；$0\leqslant t\leqslant 1$ 是单位维纳过程，且对每个 $s\in[0,\ 1]$而言，前向增量$\{I_t-I_s;\ s\leqslant t\leqslant 1\}$和 X_0^s 相互独立。此外，上述假设可确保 $P_1\ll P_0$ 成立，且有[⊖]

$$\log\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(X_0^1)=\int_0^1 \hat{S}_t\mathrm{d}X_t-\frac{1}{2}\int_0^1(\hat{S}_t)^2\mathrm{d}t \tag{6.4.77}$$

注意，并没有假设信号和噪声在 H_1 下相互独立，而仅要求噪声的前向增量与信号相互独立。因此，该模型可用于 S_t 依赖于过去观测值 X_0^t 的情形。例如，S_t 可能由观测值通过某

⊖ 这需要利用 Ito 关于 Lebesgue-Stieltjes(而非 Riemann-Stieltjes)积分的定义，并在定义中将均方收敛松弛为依概率收敛。在 Wong 和 Hajek(1985)中对此有详细讨论。

种反馈机制产生。

正是因为新息过程具有维纳统计特性，式(6.4.77)才得以成立。与拥有独立的信号和噪声的高斯情况的证明(命题6.4.6)过程类似，在式(6.4.74)的假设下，可证明，$\{I_t;0\leqslant t\leqslant 1\}$满足

$$E\{I_t-I_s\mid X_0^s\}=0,\quad 0\leqslant s\leqslant t\leqslant 1 \tag{6.4.78a}$$

且

$$E\{(I_t-I_s)^2\mid X_0^s\}=t-s,\quad 0\leqslant s\leqslant t\leqslant 1 \tag{6.4.78b}$$

一方面，维纳过程以概率1可进行连续采样；另一方面，条件式(6.4.75a)能保证$\int_0^t S_u\mathrm{d}u$和$\int_0^t \hat{S}_u\mathrm{d}u$也以概率1具有连续采样路径。所以，新息过程$\left(W_t+\int_0^t S_u\mathrm{d}u-\int_0^t \hat{S}_u\mathrm{d}u\right)$以概率1具有连续采样路径。称满足式(6.4.78a)的随机过程$\{X_0^s;0\leqslant s\leqslant 1\}$为鞅过程。显然，对每个$s\in[0,1]$，满足式(6.4.78b)的任意连续鞅过程是增量$\{I_t-I_s;s\leqslant t\leqslant 1\}$独立于$X_0^s$的维纳过程。据此，同样能得到新息过程的表达式[⊖]。

在H_1假设下，$E_1\{S_t\mid X_0^t\}$是给定X_0^s后S_t的条件均值估计，与高斯例子相同，式(6.4.77)的似然比表达式也可解释为估计-相关器。此外，式(6.4.77)对信号的约束更弱，是白噪声下检测信号的似然比的更一般表示。很多研究者基于该表达式，在对信号施加不同的约束条件后，得到了不同的估计-相关器结构，包括Duncan(1968，1970)、Kailath(1969，1971)和Stratonovich及Sosulin(1965)。Stratonovich和Sosulin采用了Stratonovich随机积分定义推广了式(6.4.77)，Schweppe(1965)推导检测了高斯信号的似然比。

作为例子，引用Kailath在1969年发表的研究成果，它可用来推导除白噪声过程之外的似然比。例如，假设有一对假设

$$\begin{aligned} H_0:Y_t&=N_t, & 0\leqslant t\leqslant 1\\ H_1:Y_t&=N_t+S_t, & 0\leqslant t\leqslant 1 \end{aligned} \tag{6.4.79}$$

其中，$\{N_t;0\leqslant t\leqslant 1\}$是具有如下形式的噪声过程

$$N_t=N_t^w+N_t^c,\quad 0\leqslant t\leqslant 1 \tag{6.4.80}$$

其中，$\{N_t^w;0\leqslant t\leqslant 1\}$是单位白噪声，噪声过程$\{N_t^c;0\leqslant t\leqslant 1\}$满足式(6.4.75)。引入假设$H:Y_t=N_t^w$，$0\leqslant t\leqslant 1$，并用与该假设相关的概率$P$作为中间量，可得到似然比$\mathrm{d}P_1/\mathrm{d}P_0$的计算式为

$$\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(X_0^1)=\frac{\mathrm{d}P_1}{\mathrm{d}P}(X_0^1)\Big/\frac{\mathrm{d}P_0}{\mathrm{d}P}(X_0^1)=\frac{\exp\left\{\int_0^1(\hat{S}_t+\hat{N}_t^1)\mathrm{d}X_t-\frac{1}{2}\int_0^1(\hat{S}_t+\hat{N}_t^1)^2\mathrm{d}t\right\}}{\exp\left\{\int_0^1(\hat{N}_t^0)\mathrm{d}X_t-\frac{1}{2}\int_0^1(\hat{N}_t^0)^2\mathrm{d}t\right\}} \tag{6.4.81}$$

其中，$\hat{N}_t^j=E_j\{N_t^c\mid X_0^t\}$，$\hat{S}_t=E_1\{S_t\mid X_0^t\}$，$X_t=\int_0^t Y_u\mathrm{d}u$，$0\leqslant t\leqslant 1$。

⊖ 这就是著名的Levy-Doob定理[参见Doob(1953)]。事实上，估计-相关器的有效性也表明新息具有维纳性质。这是Girsanov定理的一个特例[参见Girsanov(1960)和Lipster(1977)]。因此，本质上，估计-相关器和新息的性质等价。

最后需要明确：检测非高斯信号的估计器-相关器结构应更多地理解为似然比的表达形式。这是因为依据$\{\hat{S}_t;\ 0\leqslant t\leqslant 1\}$计算 dP_1/dP_0 通常非常困难。换句话说，在估计-相关器结构中求解估计量的难度不低于检测。不过，对于高斯情况，通过 Karhunen-Loéve 展开可相对容易地计算出估计量$\hat{S}_t$。在接下来的章节中，会继续讨论连续时间信号的估计问题。

6.5 习题

1. 考虑检测问题

$$H_0:Y_t = N_t,\qquad 0\leqslant t\leqslant 1$$
$$H_1:Y_t = N_t+\Theta s_t,\quad 0\leqslant t\leqslant 1$$

其中，$\{N_t:0\leqslant t\leqslant T\}$是一个零均值、均方连续的高斯过程，并且$\{s_t;\ 0\leqslant t\leqslant T\}$是一个未知信号。参数$\Theta$以等概率等于$+1$和$-1$并且与噪声独立。试求似然比和 ROC 的表达式。

2. 考虑如下假设对

$$H_0:Y_t = N_t+kJ_t+E_0\sin(\omega_0 t),\quad 0\leqslant t\leqslant T$$
$$H_1:Y_t = N_t+kJ_t+E_0\cos(\omega_0 t),\quad 0\leqslant t\leqslant T$$

其中，E_0 和 ω_0 是已知的正数；$T=2\pi/\omega_0$；$\{N_t,\ 0\leqslant t\leqslant T\}$是一个频谱高度为 $N_0/2$ 的零均值白高斯过程；$\{J_t,\ 0\leqslant t\leqslant T\}$是一个与$\{N_t,\ 0\leqslant t\leqslant T\}$独立的零均值高斯随机过程，自协方差函数为

$$C_J(t,s) = \cos(\omega_0\,|t-s|),\quad 0\leqslant t,s\leqslant T$$

假设 H_0 和 H_1 是等可能的。

(a) 假设参数 k 是已知的，推导出最小误差概率检测器并得出相应的概率或者误差。

(b) 假如一个检测器在 $k=0$ 下与(a)相同，但是实际值 $k=k_0>0$。求出 $k=k_0>0$ 时利用 $k=0$ 检测器推出的误差概率。

3. 考虑假设检验问题

$$H_0:Y_t= N_t,\quad 0\leqslant t\leqslant T$$
$$H_1:Y_t= N_t+\sum_{i=0}^{M-1}a_i p(t-iT_c)\cos(\omega_0 t+\Theta),\quad 0\leqslant t\leqslant T$$

其中，$\{N_t:0\leqslant t\leqslant T\}$是一个白高斯噪声过程，频谱高度为 $N_0/2$；序列 a_0，a_1，…，a_{M-1}是一个已知信号序列；相位Θ和噪声独立并且均匀分布在$[0,\ 2\pi]$上；函数 $p(t)$为

$$p(t) = \begin{cases}1, & 0\leqslant t\leqslant T_c\\ 0, & \text{其他}\end{cases}$$

假设 T_c 是正弦函数的整数个周期且 $T=mT_c$。

(a) 求出 H_0 和 H_1 之间的检验似然比，画出接收结构的框图。

(b) 求出虚警概率 α 的门限值。

4. 证明式(6.2.19)的正确性(已知这个结果为勒贝格分解)。

5. 考虑假设检验问题

$$H_0:Y_t = W_t,\qquad 0\leqslant t\leqslant T$$
$$H_1:Y_t = \mu t+W_t,\quad 0\leqslant t\leqslant T$$

其中$\{W_t,\ 0\leqslant t\leqslant T\}$是一个标准(单位)维纳过程。求出似然比。

6. 考虑检测问题

$$H_0:Y_t = N_t,\qquad 0\leqslant t\leqslant 1$$
$$H_1:Y_t = N_t+t^2,\quad 0\leqslant t\leqslant 1$$

其中$\{N_t,\ 0\leqslant t\leqslant T\}$是一个零均值高斯随机过程，自协方差函数为

$$C_N(t,u) = \begin{cases}t^2, & 0\leqslant t\leqslant u\leqslant 1\\ u^2, & 0\leqslant u\leqslant t\leqslant 1\end{cases}$$

证明 $P_0\equiv P_1$ 并求出 dP_1/dP_0。

7. 考虑检测问题

$$H_0:Y_t = N_t,\qquad 0\leqslant t\leqslant T$$
$$H_1:Y_t = N_t+m_0,\quad 0\leqslant t\leqslant T$$

其中，m_0 是一个正的常数，$\{N_t:0\leqslant t\leqslant T\}$是零均值和高斯的。

(a) 假设 $C_N(t,\ u)=N_0 e^{-\alpha|t-u|}$，$0\leqslant t,\ u\leqslant T$，其中 $\alpha>0$ 是固定的。证明

$$\log\frac{dP_1}{dP_0}(Y_0^T) = \frac{m_0}{2N_0}\left[Y_0+Y_T-m_0+\alpha\int_0^T\left(Y_t-\frac{m_0}{2}\right)dt\right]$$

(b) 假设

$$C_N(t,u) = \begin{cases}N_0(1-|t-u|), & |t-u|\leqslant 1\\ 0, & |t-u|>1\end{cases}$$

假设 $T<1$，证明

$$\log\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(Y_0^T)=\frac{m_0}{N_0(2-T)}(Y_0+Y_T-m_0)$$

8. 考虑式(6.3.47)中的模型。假设基于信号模型 $m_t=\int_0^t s_u\mathrm{d}u$，$0\leqslant t\leqslant T$ 设计一个尼曼-皮尔逊检测器，但是实际信号由 $\int_0^t r_u\mathrm{d}u$，$0\leqslant t\leqslant T$ 给出，其中$\{r_u;\ 0\leqslant u\leqslant T\}$是一个确定信号。
   (a) 求出在实际信号情况下所设计检测器的 P_F 和 P_D。
   (b) 假设 $\int_0^T r_t^2\mathrm{d}t=\int_0^T s_t^2\mathrm{d}t=P>0$，简述由(a)得出的是 $\int_0^T r_t s_t\mathrm{d}t$ 的函数的检测概率。
9. 考虑在高斯白噪声情况下的模型(6.3.1)。求出在约束条件 $\max\left\{\int_0^T[s_t^0]^2\mathrm{d}t,\int_0^T[s_t^1]^2\mathrm{d}t\right\}\leqslant P$ 下的最优信号，其中 $P>0$。
10. 计算在式(6.4.24)最优检测统计下的式(6.4.31)的偏差率。
11. 证明式(6.4.56)的有效性。

第7章

连续时间信号估计

7.1 引言

第 6 章讨论了时域连续信号的检测问题，在这一章中将继续讨论时域连续信号的估计问题。本章主要解决以下三个基本问题：(1)确定信号参数估计：在加性高斯白噪声下估计波形已知或部分参数未知的确定信号的参数；(2)线性/高斯估计：假设信号和噪声均为高斯过程，或者仅考虑线性估计器的情况；(3)非线性滤波：在加性高斯白噪声中，估计由非线性微分方程产生的非高斯随机信号。根据第 6 章的证明，其他高斯噪声模型都可以转化为白噪声模型，因此本章主要考虑高斯白噪声的情况，相应的结论也具有一般性。

7.2 信号参数估计

首先考虑含噪声情况下连续信号的参数估计问题。类似于离散信号，观测模型如下

$$Y_t = s_t(\theta) + N_t,\quad 0 \leqslant t \leqslant T \tag{7.2.1}$$

其中波形$\{s_t(\theta);\ 0\leqslant t\leqslant T\}$与 θ 的函数关系式已知，$\{N_t;\ 0\leqslant t\leqslant T\}$是功率谱幅度为 $N_0/2$ 的高斯白噪声。对 Y_t 求积分可得

$$X_t = \int_0^t s_u(\theta)\,\mathrm{d}u + W_t,\quad 0 \leqslant t \leqslant T \tag{7.2.2}$$

其中$\{W_t;\ 0\leqslant t\leqslant T\}$表示相应的维纳过程。

为使用与第 4 章相同的方法来估计式(7.2.2)中的 θ，首先需在观测空间$(\Gamma,\ \zeta)$中指定概率密度族$\{p_\theta;\ \theta\in\Lambda\}$，而为了获得这一概率密度族，进一步要求存在某个测度 μ 使得在概率密度族$\{P_\theta;\ \theta\in\Lambda\}$上的其他测度绝对连续。由例 6.3.1 可知，如果 $s_t(\theta)$在$[0,\ T]$是平方可积的，不存在信号时$(X_t=W_t,\ 0\leqslant t\leqslant T)$的概率测度(称之为维纳测度)能确保 P_θ 绝对连续。假设该测度由 μ 表示，则密度 $P_\theta=\mathrm{d}P_\theta/\mathrm{d}\mu$ 为式(6.3.65)所描述的似然比，即

$$p_\theta(X_0^T) = \exp\left\{\frac{2}{N_0}\int_0^T s_t(\theta)\,\mathrm{d}X_t - \frac{1}{N_0}\int_0^T s_t^2(\theta)\,\mathrm{d}t\right\} \tag{7.2.3}$$

假设对于每一个 $\theta\in\Lambda$，$\{s_t(\theta);\ 0\leqslant t\leqslant T\}$都平方可积，则式(7.2.3)给出了 X_0^T 的一个概率密度族。

在得到集合$\{p_\theta;\ \theta\in\Lambda\}$后，就可以使用第 4 章的方法来估计 θ 的值。如果在随机变量 Θ 中，θ 的先验概率密度函数为 ω，那么根据后验概率密度函数可得到贝叶斯估计为

$$\omega(\theta\,|\,X_0^T) = \frac{p_\theta(X_0^T)\omega(\theta)}{\int_\Lambda p_\theta(X_0^T)\omega(\theta)\,\mathrm{d}\theta} \tag{7.2.4}$$

类似地，其他非贝叶斯方法，如最小方差无偏估计和最大似然估计，均可直接应用到模型

$\{p_\theta;\ \theta\in\Lambda\}$上。

为了详细讨论模型(7.2.2)下最大似然估计的渐近特性，假设参数集 Λ 是 $\mathbf{R}$ 上的一个开放区间，并且像离散信号一样可扩展至参数向量的形式。

首先，这一模型的 Fisher 信息可由下式给出：

$$\begin{aligned}I_\theta &\triangleq E_\theta\left\{\left[\frac{\partial}{\partial\theta}\log p_\theta(X_0^T)\right]^2\right\}=E_\theta\left\{\left[\frac{2}{N_0}\frac{\partial}{\partial\theta}\left[\int_0^T s_t(\theta)\mathrm{d}X_t-\frac{1}{2}\int_0^T s_t^2(\theta)\mathrm{d}t\right]\right]^2\right\}\\&=\frac{4}{N_0^2}E_\theta\left\{\left[\int_0^T s_t'(\theta)(\mathrm{d}X_t-s_t(\theta)\mathrm{d}t)\right]^2\right\}=\frac{4}{N_0^2}E_\theta\left\{\left[\int_0^T s_t'(\theta)\mathrm{d}W_t\right]^2\right\}=\frac{2}{N_0}\int_0^T[s_t'(\theta)]^2\mathrm{d}t\end{aligned}\tag{7.2.5}$$

上式中，在第三个等式中用 $s_t'(\theta)=\partial s_t(\theta)/\partial\theta$ 表示对 θ 求偏微分且假设其能够移到积分括号里面，第四个等式中使用了式(7.2.2)，最后一个等式的推导使用了命题 6.4.1。式(7.2.5)除以 T 的比值可用来表示 X_0^T 产生的关于 θ 的信息的快慢，假设在 $T\to\infty$时该比值趋于一个非零的有限值 i_θ，即对所有 $\theta\in\Lambda$ 有下式成立

$$0<i_\theta\triangleq\lim_{T\to\infty}\frac{2}{N_0T}\int_0^T[s_t'(\theta)]^2\mathrm{d}t<\infty\tag{7.2.6}$$

该假设意味着通过观测值得到参数信息的过程是渐近线性的。注意，在 4.3 节中处理的独立同分布离散问题中，通过观测值获得参数信息的过程是严格线性的，这是因为基于 $Y_1,\ \cdots,\ Y_n$的 I_θ 具有 ni_θ 的形式。

具有如(7.2.3)形式的密度族的似然比方程为

$$\frac{\partial}{\partial\theta}\left[\frac{2}{N_0}\int_0^T s_t(\theta)\mathrm{d}X_t-\frac{1}{N_0}\int_0^T s_t^2(\theta)\mathrm{d}t\right]\bigg|_{\theta=\hat{\theta}_{ML}(X_0^T)}=0\tag{7.2.7}$$

再次假设上式中的偏微分可移到积分里面，则上式等价于

$$J_\theta(X_0^T)\big|_{\theta=\hat{\theta}_{ML}(X_0^T)}=0\tag{7.2.8}$$

其中

$$J_\theta(X_0^T)\triangleq\frac{1}{T}\int_0^T s_t'(\theta)[\mathrm{d}X_t-s_t(\theta)\mathrm{d}t]\tag{7.2.9}$$

上式中 $s_t'(\theta)$ 的含义同前，且为了简便除以了 $2T/N_0$。用 θ_0 表示真实参数并联合式(7.2.2)可进一步得到

$$\begin{aligned}J_\theta(X_0^T)&=\frac{1}{T}\int_0^T s_t'(\theta)\{\mathrm{d}W_t+[s_t(\theta_0)-s_t(\theta)]\mathrm{d}t\}\\&=\frac{1}{T}\int_0^T s_t'(\theta)\mathrm{d}W_t+\frac{1}{T}\int_0^T s_t'(\theta)[s_t(\theta_0)-s_t(\theta)]\mathrm{d}t\end{aligned}\tag{7.2.10}$$

根据命题 6.4.1，上式右边的第一部分服从$\mathcal{N}(0,\ I_\theta/T^2)$分布，其中 I_θ 根据式(7.2.5)计算。由于 $I_\theta/T\to i_\theta<\infty$，第一部分的方差在 $T\to\infty$时趋向于零，因此当 $T\to\infty$时，这一部分均方意义上(依概率)收敛。假设式(7.2.10)中第二部分收敛，则在 P_{θ_0} 下有

$$J_\theta(X_0^T)\xrightarrow{\text{依概率}}\lim_{T\to\infty}\frac{1}{T}\int_0^T s_t'(\theta)[s_t(\theta_0)-s_t(\theta)]\mathrm{d}t\triangleq J(\theta;\theta_0)\tag{7.2.11}$$

其中函数 $J(\theta;\ \theta_0)$在 $\theta=\theta_0$ 处存在一个根。该渐近形式与离散信号下的情况类似：如果 $J(\theta;\ \theta_0)$在 $\theta=\theta_0$ 处存在唯一解，便可认为$\hat{\theta}_{ML}(X_0^T)$在一定条件下收敛至准确值。因此，类似于命题 4.4.3，有下面的命题成立。

命题 7.2.1　一致性

假设参数 θ_0 为真值，且 $J(\theta;\ \theta_0)$ 在 $\theta=\theta_0$ 处有一个孤立解，即函数 $J(\theta;\ \theta_0)$ 在 θ_0 点改变符号，并且在 θ_0 的邻域内，函数 $J(X_0^T)$ 都是 θ 的连续函数。这样，在 $T\to\infty$ 时，$J_\theta(X_0^T)=0$ 便有一组解序列依概率收敛到 θ_0。

注意　该命题的证明过程与命题 4.4.3 类似，留作课后练习。命题 7.2.1 的一个推论是：如果对每一个 T，函数 $J_\theta(X_0^T)$ 都有唯一的根，则这个根序列是一致的。

接下来讨论似然比方程解的渐近正态性。事实上，与 4.4.4 类似，可证明似然方程的解具有渐近正态性。

命题 7.2.2　渐近正态性

假设 $\{\hat{\theta}_T(X_0^T);\ T\geqslant 0\}$ 是方程 $J_\theta(X_0^T)=0$ 的一致解序列，进一步假设 $\partial^2 s_t(\theta)/\partial\theta^2$ 和 $\partial^3 s_t(\theta)/\partial\theta^3$ 存在且在 $T\geqslant 0$ 和 $\theta\in\Lambda$ 时一致有界，则有

$$\sqrt{T}[\hat{\theta}_T(X_0^T)-\theta_0]\to\mathcal{N}(0,1/i_{\theta_0})$$

其中 θ_0 是参数的真实值，i_θ 由式(7.2.6)定义。

注意　该命题说明在相同情况下，模型(7.2.2)中的最大似然估计值和离散情况类似，均是渐近有效的。

以上渐近特性的证明中，假设 $T\to\infty$ 且 $0<i_\theta<\infty$，下面考虑更加一般的情况。设 I_θ 没有上限(上面的分析中假设 $I_\theta\sim i_\theta T$，因此是一种特例)，除去假设 $0<i_\theta<\infty$ 并重写渐近方程(7.2.9)

$$\left.\frac{\int_0^T s'_t(\theta)[\mathrm{d}X_t-s_t(\theta)\mathrm{d}t]}{\int_0^T[s'_t(\theta)]^2\mathrm{d}t}\right|_{\theta=\hat{\theta}_{ML}(X_0^T)}=0 \tag{7.2.12}$$

类似于式(7.2.10)中的展开方式，式(7.2.12)的左边可写为

$$\frac{\int_0^T s'_t(\theta)\mathrm{d}W_t}{\int_0^T[s'_t(\theta)]^2\mathrm{d}t}+\frac{\int_0^T s'_t(\theta)[s_t(\theta_0)-s_t(\theta)]\mathrm{d}t}{\int_0^T[s'_t(\theta)]^2\mathrm{d}t} \tag{7.2.13}$$

显然式(7.2.13)的第一项服从 $\mathcal{N}(0,\ 1/I_\theta)$ 分布，当 $I_\theta\to\infty$ 时这部分均方收敛于零，进而可得出在上述假设条件下，解的渐近有效性可表示为关于 θ_0 的函数，即

$$\lim_{I_\theta\to\infty}\left[\frac{\frac{1}{N_0}\int_0^T s'_t(\theta)[s_t(\theta_0)-s_t(\theta)]\mathrm{d}t}{I_\theta}\right]\triangleq\tilde{J}(\theta;\theta_0) \tag{7.2.14}$$

和命题 7.2.1 的结论一样，在对 $\tilde{J}$ 进行适当假设后，随着 $I_\theta\to\infty$，该似然方程的解具有一致性。

而当 $I_{\theta_0}\to\infty$ 时，证明似然方程连续解的渐近正态性可采用类似于命题 7.2.2 的方法。特别地，在与命题 7.2.2 相同的正则条件下，$I_{\theta_0}\to\infty$ 时，似然方程的任意连续一致解依概率满足

$$I_{\theta_0}^{1/2}[\hat{\theta}_T(X_0^T)-\theta_0]\to\mathcal{N}(0,1) \tag{7.2.15}$$

注意当 $T\to\infty$、$N_0\to 0$ 和 $\int_0^T[s'_t(\theta)]^2\mathrm{d}t\to\infty$ 中的任意一个条件成立时，I_{θ_0} 都可无限制

地增加。除了 $T\to\infty$ 之外，考虑 $N_0\to 0$（其他参数固定）可用于分析估计器在高信噪比下的性能。此外，虽然似然方程的解并不依赖于 N_0，但是只要其他参数不变，随着 $N_0\to 0$，就可以得到固定的估计值，其分布随 N_0 变化，下面结合具体例子来说明。

例 7.2.1 到达时间估计

在许多应用中，需要确定已知波形信号到达接收器的具体时间。最典型的应用是雷达测距：通过发射一个已知形式的信号，经过目标反射后被天线所接收，由返回信号的到达时间便可以算出信号来回的传播时间，也就可以推算出目标与雷达的距离。

估计到达时间的参数化信号模型如下

$$s_t(\theta) = s(t-\theta),\quad 0\leqslant t\leqslant T \tag{7.2.16}$$

这里信号波形 $\{s(t);\ t\in\mathbf{R}\}$ 已知且在 $t<0$ 时为零，$\theta\geqslant 0$ 是信号的到达时间。该信号的 Fisher 信息是

$$I_\theta = \frac{2}{N_0}\int_0^T [s'(t-\theta)]^2\,\mathrm{d}t \tag{7.2.17}$$

其中 $s'(t)=\mathrm{d}s(t)/\mathrm{d}t$。这类信号通常满足 $\int_0^\infty [s'(t-\theta)]^2\,\mathrm{d}t<\infty$，此时渐近信息率 i_q 就等于零，因此不能直接使用 $T\to\infty$ 下的渐近计算来得到最大似然估计。但是在高信噪比下，可借助 $n_0\to 0$ 得到到达时间的渐近估计值。

在 $n_0\to 0$ 时，似然方程的一致解具有渐近方差 I_θ^{-1}，在 $T\to\infty$ 时有

$$I_\theta = \frac{2}{N_0}\int_0^\infty [s'(t)]^2\,\mathrm{d}t = \frac{1}{\pi N_0}\int_{-\infty}^\infty \omega^2\,|S(\omega)|^2\,\mathrm{d}\omega \tag{7.2.18}$$

其中 $S(\omega)$ 是 $s(t)$ 的傅里叶变换（第二个等式基于帕斯瓦尔能量定理）。渐近误差的方差为

$$\mathrm{var}_\theta(\hat{\theta}_\infty) \sim \left[\frac{2E_s}{N_0}\beta^2\right]^{-1} \tag{7.2.19}$$

其中 $E_s \triangleq \int_0^\infty s^2(t)\,\mathrm{d}t \equiv \int_{-\infty}^\infty |S(\omega)|^2\,\mathrm{d}\omega/2\pi$ 为信号能量，β 为信号带宽的均方根，定义如下

$$\beta \triangleq \left[\frac{1}{2\pi}\int_{-\infty}^\infty \omega^2\,|S(\omega)|^2\,\mathrm{d}\omega\Big/\frac{1}{2\pi}\int_{-\infty}^\infty |S(\omega)|^2\,\mathrm{d}\omega\right]^{1/2} \tag{7.2.20}$$

通过增大信噪比 $2E_s/N_0$ 或信号带宽，可以提高到达时间的估计精度。增大信号带宽改善到达时间的估计精度是非常直观的，因为带宽的增加将会使得信号边缘变得更加“陡峭”，利于估计精度的提高。

最后考虑对数似然函数的情况，其信号形式可表达为

$$\log p_\theta(X_0^T) = \frac{2}{N_0}\int_0^T s(t-\theta)\,\mathrm{d}X_t - \frac{1}{N_0}\int_0^T s^2(t-\theta)\,\mathrm{d}t \tag{7.2.21}$$

假设信号持续时间 $D<T$ 且 θ 在区间 $[0,\ T-D]$ 中，此时等式右边的第二项与变量 θ 无关，因此最大似然估计即为使得 $\int_0^\infty s(t-\theta)\,\mathrm{d}X_t$ 最大的 θ 值。从应用的角度出发，最大似然估计方法需计算每个 θ 对应的 $\int_0^\infty s(t-\theta)Y_t\,\mathrm{d}X_t$（$Y_t$ 可视为由 X_t 派生的信号）并选择输出峰值时刻所对应的 θ 作为其估计值。最大似然估计的计算过程看似繁琐，但由于 Y_t 是时间有限信号，$\int_0^\infty s(t-\theta)Y_t\,\mathrm{d}X_t$ 实际上是脉冲响应 $h(t)=s(-t)$ 的线性时不变滤波器在 θ 时刻的输

出，因此将 Y_t 通过滤波器 $h(t)$ 便可以简单地计算出 θ 的最大似然估计。◀

7.3 线性/高斯估计

7.3.1 白噪声中的信号估计问题

下面讨论在加性独立高斯白噪声下的均方连续高斯信号估计问题。信号模型如下

$$Y_t = S_t + N_t, \quad 0 \leqslant t \leqslant T \tag{7.3.1}$$

其中，$\{S_t, 0\leqslant t\leqslant T\}$ 是具有连续自协方差函数 $C_s(t, u)$ 的零均值高斯信号，$\{N_t; 0\leqslant t\leqslant T\}$ 是功率谱幅度为 $N_0/2$ 的高斯白噪声（与信号相互独立）。类似于前面的做法，首先对 $\{Y_t; 0\leqslant t\leqslant T\}$ 积分得到下面模型

$$X_t = \int_0^t S_u \mathrm{d}u + W_t \quad 0 \leqslant t \leqslant T \tag{7.3.2}$$

其中 $\{W_t; 0\leqslant t\leqslant T\}$ 是维纳过程。

对于 $[0, T]$ 中的任意固定值 t，面临的主要问题是如何利用观测值 $X_0^s=\{X_u; 0\leqslant u\leqslant s\}$ 获得 S_t 在最小均方误差(MMSE)意义上的最优估计量。为了得到最优估计量，首先考虑在指定 X_0^s 下得到 S_t 的最优线性估计量，然后再证明它是全局最优的。

为获得最优的线性估计值，选定 $t\in[0, T]$ 并考虑如下线性估计

$$\hat{S}_t = \int_0^s h(t,v)\mathrm{d}X_v \tag{7.3.3}$$

其中 $h(t, v)$ 是 v 的连续函数。根据 5.3 节的相关内容，最佳线性估计值的充要条件是估计误差 $\hat{S}_t - S_t$ 与观测值正交，即

$$E\{(\hat{S}_t - S_t)X_u\} = 0, \quad 0 \leqslant u \leqslant s \tag{7.3.4}$$

代入式(7.3.2)和式(7.3.3)，则式(7.3.4)可写为

$$\begin{aligned} &E\left\{W_u\int_0^s h(t,v)\mathrm{d}W_v\right\}+E\left\{\int_0^u S_q\mathrm{d}q\int_0^s h(t,v)\mathrm{d}W_v\right\} \\ &+E\left\{W_u\int_0^s h(t,v)S_v\mathrm{d}v\right\}+E\left\{\int_0^u S_q\mathrm{d}q\int_0^s h(t,v)S_v\mathrm{d}v\right\} \\ &=E\{S_tW_u\}+E\left\{S_t\int_0^u S_q\mathrm{d}q\right\}, \quad 0 \leqslant u \leqslant s \end{aligned} \tag{7.3.5}$$

由于信号和噪声相互独立，且根据命题 6.4.1 中的性质 6)，式(7.3.5)可以写为[⊖]

$$\frac{N_0}{2}\int_0^u h(t,v)\mathrm{d}v + \int_0^u\int_0^s h(t,v)C_s(v,q)\mathrm{d}v\mathrm{d}q = \int_0^u C_s(t,q)\mathrm{d}q, \quad 0 \leqslant u \leqslant s \tag{7.3.6}$$

在式(7.3.6)两边同时对 u 求偏导，可以得到如下结论：当且仅当 h 满足式(7.3.7)时，由式(7.3.3)给出的 $\hat{S}_t$ 是最佳估计

$$\frac{N_0}{2}h(t,u) + \int_0^s h(t,v)C_s(v,u)\mathrm{d}v = C_s(t,u), \quad 0 \leqslant u \leqslant s \tag{7.3.7}$$

一旦给定 t，式(7.3.7)是第二类 Fredholm 积分方程。可以证明，如果假设 C_s 为连续函数，那么式(7.3.7)存在关于参数 u 连续的唯一解 $h(t, u)$[详细参见 Lovitt(1950)关于

⊖ 积分与期望的顺序可以互换。

Fredholm 方程的讨论]。

至此，已经证明：给定 X_0^s 后，h 满足式(7.3.7)时，式(7.3.3)中的$\hat{S}_t$ 是 MMSE 线性估计。注意，式(7.3.7)是连续时间 Wiener-Hopf 方程，并且此前已经讨论了它在高斯白噪声下检测信号的两个特例：$s=t$ 和 $s=T$。

下面将证明在给定 X_0^s 下，$\hat{S}_t$ 不仅是 S_t 的最佳线性估计量，也是 S_t 所有可能估计的最优值。注意，一般意义上的估计器可视为随机变量的集合，这些随机变量是 X_0^s 的函数且二阶矩有限。将所有可能的估计值用$\mathcal{G}_0^s$ 来表示后，正交性原理非常直观。

命题 7.3.1 正交性

$$E\{(Y-S_t)^2\} = \min_{Z\in\mathcal{G}_0^s} E\{(Z-S_t)^2\} \tag{7.3.8}$$

对所有 $Z\in\mathcal{G}_0^s$，当且仅当下面的条件满足时式(7.3.8)成立

$$E\{(Y-S_t)Z\} = 0 \tag{7.3.9}$$

如果 Y_1 和 Y_2 对所有的 $Z\in\mathcal{G}_0^s$ 均满足式(7.3.9)，则 Y_1 和 Y_2 以概率 1 相等。

根据第 4 章的分析，$Y=E\{S_t|X_0^s\}$是式(7.3.8)的解，因此正交条件(7.3.9)就是在指定 X_0^s 下，S_t 的条件均值所需满足的特性。

就式(7.3.3)中的$\hat{S}_t$ 而言，$\hat{S}_t-S_t$ 与 X_0^s 不相关，且由于($\hat{S}_t-S_t$)与 X_0^s 服从联合高斯分布(因为($\hat{S}_t-S_t$)是高斯变量(S_t，X_0^s)的线性组合)，故($\hat{S}_t-S_t$)独立于 X_0^s，这意味着对 $Z\in\mathcal{G}_0^s$ 有

$$E\{(\hat{S}_t-S_t)Z\} = E\{(\hat{S}_t-S_t)\}E\{Z\}$$

由于 $E\{\hat{S}_t\}=E\{S_t\}=0$ 时式(7.3.9)成立，所以，在高斯模型下的线性估计量也是全局最优估计值。

但是应当注意，在证明过程中并没有限定高斯假设，该模型仅需满足如下条件：$E\{W_tW_u\}=(N_0/2)\min\{t, u\}$，$\mathrm{cov}(S_t, S_u)=C_s(t, u)$，$\mathrm{cov}(S_tW_u)=0$，$E\{S_t\}=E\{W_t\}=0$，$0\leqslant u, t\leqslant T$。因此噪声只要满足上述特性，无论是高斯的或非高斯的信号，式(7.3.3)均为最优的线性估计量。

只要信号与噪声服从联合高斯分布，就可以证明线性估计量的全局最优性。而在此条件下，上述所有的结果都可以直接扩展到信号和噪声相关的情况。类似地，在噪声由均方连续随机变量和白高斯两个部分构成时，上述结论也成立。

7.3.2 线性新息过程

观测模型为

$$X_t = \int_0^t S_u\mathrm{d}u + W_t,\quad 0\leqslant t\leqslant T \tag{7.3.10}$$

其中对所有 $0\leqslant t, u\leqslant t$，假设下列条件成立：$E\{W_tW_u\}=\min\{t, u\}$，$C_s(t, u)$连续且 $E\{S_t\}=E\{W_t\}=0$。则对于任意的 $t\in[0, T]$，如果设$\hat{S}_t$ 为观测数据 X_0^s 下 S_t 的最佳线性估计量[即在 $s=t$ 和 $N_0/2=1$ 下的式(7.3.3)]。根据 6.4 节中的结论，以下过程的均值为零

$$I_t = X_t - \int_0^t \hat{S}_u\mathrm{d}u,\quad 0\leqslant t\leqslant T \tag{7.3.11}$$

且对于 $0\leqslant u\leqslant t\leqslant T$ 的协方差 $C_I(t,\ u)=\min\{t,\ u\}$ 和 I_t-I_u 不依赖于 X_0^u。此外，如果式(7.3.10)是高斯过程，那么 I_0^T 同样也是高斯的(维纳过程)，且 I_t-I_u 和 X_0^u 独立。为了与已定义的新息过程 $\{X_t-\int_0^t E\{S_u\mid X_0^u\}\mathrm{d}u\ ,\ 0\leqslant t\leqslant T\}$ 相区分，称 I_t 为线性新息过程，这两个过程在高斯情况下一致。

由于 I_0^T 中含有白噪声的协方差，获得线性新息的过程在本质上等价于对 X_0^T 进行线性白化操作，并且不同于6.4节讨论的预白化处理，I_0^T 是由 X_0^T 因果产生的。然而，由于不能证明 X_0^T 可由 I_0^T 推导获得，因此不能确定线性新息过程是否在任何情况下均等价于最初的观测结果。接下来，重点推导线性新息过程与原始观测的等价性。

已知 X_0^T 与 I_0^T 的联系如下式所示

$$X_t=I_t+\int_0^t\hat{S}_u\mathrm{d}u,\quad 0\leqslant t\leqslant T \tag{7.3.12}$$

如果可以通过某种方法由 I_0^u 得到 $\hat{S}_u$，则 X_0^T 就可以通过式(7.3.12)构造出来。例如，假设对于任意的 $t\in[0,\ T]$，用 $\widetilde{S}_t$ 来表示给定 I_0^t 下 S_t 的MMSE线性估计值后，如果 $\hat{S}_t=\widetilde{S}_t$ 成立，则 X_0^T 就可以计算出来。

利用正交性原理和 I_0^t 的定义，$\widetilde{S}_t$ 可表示为

$$\widetilde{S}_t=\int_0^t g(t,u)\mathrm{d}I_u,\quad 0\leqslant t\leqslant T \tag{7.3.13}$$

其中函数 g 定义为

$$g(t,u)=C_s(t,u)-\int_0^u C_s(t,u)h(u,v)\mathrm{d}v,\quad 0\leqslant t\leqslant T \tag{7.3.14}$$

$h(u,\ v)$ 是求解 $\hat{S}_u$ 的冲激响应 $\left[\text{即}\hat{S}_u=\int_0^u h(u,v)\mathrm{d}X_v\right]$。用 $X_u-\int_0^u\hat{S}_v\mathrm{d}v$ 代替 I_u，得到如下表达

$$\begin{aligned}\hat{S}_t&=\int_0^t g(t,u)\mathrm{d}X_u-\int_0^t g(t,u)\ \hat{S}_u\mathrm{d}u=\int_0^t g(t,u)\mathrm{d}X_u-\int_0^t g(t,u)\int_0^u h(u,v)\mathrm{d}X_v\mathrm{d}u\\&=\int_0^t g(t,u)\mathrm{d}X_u-\int_0^t\int_v^t g(t,u)h(u,v)\mathrm{d}u\mathrm{d}X_v\\&=\int_0^t\left[g(t,u)-\int_u^t g(t,v)h(v,u)\mathrm{d}v\right]\mathrm{d}X_u\end{aligned} \tag{7.3.15}$$

上式中第三个等式交换了 u 和 v 的积分顺序，第四个等式中改变了第二个积分中的积分变量。将式(7.3.14)代入式(7.3.15)中，整理得到

$$\widetilde{S}_t=\int_0^t f(t,u)\mathrm{d}X_u \tag{7.3.16}$$

其中

$$f(t,u)\triangleq C_s(t,u)-\int_0^t Q_t(u,v)C_s(t,v)\mathrm{d}v \tag{7.3.17}$$

且

$$Q_t(u,v)\triangleq h(u,v)+h(v,u)-\int_0^t h(\alpha,u)h(\alpha,v)\mathrm{d}\alpha,\quad 0\leqslant u,v\leqslant t \tag{7.3.18}$$

$h(x, y)$在 $t\geqslant x\geqslant y\geqslant 0$ 时等于零。

函数 $Q_t(u, v)$是和 $C_s(u, v)$，$0\leqslant u, v\leqslant t$ 有关的预解(或者称为 Fredhom 预解)。第 6 章已经证明，Q_t 是如下积分方程的解

$$C_s(u,v) = \int_0^t Q_t(u,\alpha)C_s(\alpha,v)\mathrm{d}\alpha + Q_t(u,v), \quad 0\leqslant u,v\leqslant t$$

Q_t 等价为以下积分方程的预解函数

$$y(u) = \int_0^t C_s(u,v)x(v)\mathrm{d}v + x(u), \quad 0\leqslant u\leqslant t \tag{7.3.19}$$

因为该方程的唯一连续解为

$$x(u) = y(u) - \int_0^t Q_t(u,v)y(v)\mathrm{d}v, \quad 0\leqslant u\leqslant t \tag{7.3.20}$$

根据式(7.3.17)、式(7.3.18)和式(7.3.20)，可以证明 $f(t, u)$满足

$$C_s(t,u) = \int_0^t C_s(u,v)f(t,v)\mathrm{d}v + f(t,u), \quad 0\leqslant u\leqslant t$$

上式即为求解 $h(t, u)$的 Wiener-Hopf 方程[与 $s=t$，$N_0/2=1$ 时的式(7.3.7)相同]。已知式(7.3.7)具有唯一解，可得 $f(t, u)=h(t, u)$，$0\leqslant u\leqslant t$ 且有

$$\widetilde{S}_t = \int_0^t h(t,u)\mathrm{d}X_u = \hat{S}_t$$

至此，证明了$\hat{S}_t$ 完全由 I_0^t 决定，进而 X_0^t 也能由 I_0^t 重构。更准确地说，线性新息过程 I_0^T 可由 X_0^T 的线性因果变换得到，且根据逆变换可由 X_0^T 得到 I_0^T。该变换对可以写为

$$I_t = \int_0^t \left[1 - \int_u^t h(u,v)\mathrm{d}v\right]\mathrm{d}X_u, \quad 0\leqslant t\leqslant T \tag{7.3.21}$$

以及

$$X_t = \int_0^t \left[1 + \int_u^t g(u,v)\mathrm{d}v\right]\mathrm{d}I_u, \quad 0\leqslant t\leqslant T \tag{7.3.22}$$

其中 h 和 g 在式(7.3.14)中已定义。因此 I_0^T 和 X_0^T 是一对等价观测量，进一步地，基于 I_0^T 或 X_0^T 得到的最佳线性估计量具有一致性。

线性新息过程的上述特性可以直接扩展到向量信号形式。特别地，假设 n 维观测向量可由下式给出

$$\boldsymbol{X}_t = \int_0^t \boldsymbol{S}_u\mathrm{d}u + \boldsymbol{W}_t, \quad 0\leqslant t\leqslant T \tag{7.3.23}$$

其中$\{\boldsymbol{S}_t;\ 0\leqslant t\leqslant T\}$是具有连续自协方差的零均值向量信号⊖，其协方差为

$$C_s(t,u) \triangleq E\{\boldsymbol{S}_t\boldsymbol{S}'_u\}, \quad 0\leqslant u,t\leqslant T$$

$\{\boldsymbol{W}_t;\ 0\leqslant t\leqslant T\}$是零均值向量过程，自协方差矩阵为

$$E\{\boldsymbol{W}_t\boldsymbol{W}'_u\} = \boldsymbol{I}\min\{t,u\}, \quad 0\leqslant u,t\leqslant T$$

其中 $\boldsymbol{I}$ 是 $n\times n$ 的单位矩阵。值得注意的是：向量 $\boldsymbol{W}_0^T$ 由一组不相关的随机过程组成，这些随机过程的自协方差均为 $\min\{t, u\}$。如果 $\boldsymbol{W}_0^T$ 是高斯的，那么它就是向量维纳过程，可理解为独立高斯白噪声组成的向量的积分。

⊖ 为避免混淆，代表转置的符号“T”用“·”替代。

对区间[0，T]上的所有 t 和 u 而言，假设 $\boldsymbol{S}_u$ 和 $\boldsymbol{W}_t$ 不相关，则给定 $\boldsymbol{X}_0^t$ 下 $\boldsymbol{S}_t$ 的最佳线性估计由下式给出[⊖]

$$\hat{\boldsymbol{S}}_t=\int_0^t \boldsymbol{h}(t,u)\mathrm{d}\boldsymbol{X}_u \tag{7.3.24}$$

$n\times n$ 维矩阵函数 $h(t, u)$为下面矩阵积分方程的解：

$$\boldsymbol{C}_s(t,u)=\int_0^t \boldsymbol{h}(t,v)\boldsymbol{C}_s(v,u)\mathrm{d}v+\boldsymbol{h}(t,u),\quad 0\leqslant u\leqslant t \tag{7.3.25}$$

它由下述充要条件得到，即

$$E\{(\boldsymbol{S}_t-\hat{\boldsymbol{S}}_t)\boldsymbol{X}'_u\}=0,\quad 0\leqslant u\leqslant t \tag{7.3.26}$$

其中 **0** 是 $n\times n$ 的零矩阵，并且有

$$E\left\{\left(\int_0^T \boldsymbol{A}_t\mathrm{d}\boldsymbol{W}_t\right)\left(\int_0^T \boldsymbol{B}_t\mathrm{d}\boldsymbol{W}_t\right)'\right\}=\int_0^T \boldsymbol{A}_t\boldsymbol{B}_t{}'\mathrm{d}t \tag{7.3.27}$$

其中，$\boldsymbol{A}_t$ 和 $\boldsymbol{B}_t$ 是由平方可积的元素组成的矩阵函数，且 $\boldsymbol{W}_0^T$ 具有自协方差函数 $\boldsymbol{I}\cdot\min\{t, u\}$。

类似于单个随机变量，新息过程

$$\boldsymbol{I}_t\triangleq\boldsymbol{X}_t-\int_0^t \hat{\boldsymbol{S}}_u\mathrm{d}u,\quad 0\leqslant t\leqslant T \tag{7.3.28}$$

满足如下条件

$$E\{\boldsymbol{I}_t\boldsymbol{I}'_u\}=\boldsymbol{I}\min\{t,u\},\quad 0\leqslant u,t\leqslant T \tag{7.3.29}$$

和

$$E\{(\boldsymbol{I}_t-\boldsymbol{I}_s)\boldsymbol{X}_u{}'\}=0,\quad 0\leqslant u\leqslant s\leqslant t\leqslant T \tag{7.3.30}$$

进一步地，$\boldsymbol{X}_0^T$ 可表达为 $\boldsymbol{I}_0^T$ 的函数，即

$$\boldsymbol{X}_t=\int_0^t\left[\boldsymbol{I}+\int_0^t \boldsymbol{g}(u,v)\mathrm{d}v\right]\mathrm{d}\boldsymbol{I}_u,\quad 0\leqslant t\leqslant T \tag{7.3.31}$$

其中

$$\boldsymbol{g}(t,u)=\boldsymbol{C}_s(t,u)-\int_0^u \boldsymbol{C}_s(t,v)\boldsymbol{h}'(u,v)\mathrm{d}v,\quad 0\leqslant u\leqslant t\leqslant T$$

如果信号和噪声均是高斯分布(并且是联合高斯的)，则式(7.3.24)中的 $\hat{\boldsymbol{S}}_t$ 是全局 MMSE 估计量，且在所有 $t\geqslant s\geqslant 0$ 下，维纳过程向量 $\boldsymbol{I}_0^T$ 含有不依赖于 X_0^s 的增量 $\boldsymbol{I}_t-\boldsymbol{I}_s$。

7.3.3　连续时间 Kalman-Bucy 滤波器

在第 5 章中推导了 Kalman-Bucy 滤波器的时域表达时，本节将结合信息过程来完成 Kalman-Bucy 滤波器的推导。

大多数物理系统在连续时间上的系统特性都可以在一定时间区间内表示(例如[0，T])，任意时间 $t\in$[0，T]对应的状态向量 $\boldsymbol{X}_t$ 可通过如下线性微分方程得到

$$\boldsymbol{X}_t=\boldsymbol{A}_t\boldsymbol{X}_t+\boldsymbol{B}_t\boldsymbol{U}_t,\quad 0\leqslant t\leqslant T \tag{7.3.32a}$$

该方程由随机过程$\{\boldsymbol{U}_t;\ t\in[0, T]\}$驱动。对于每一个 t，状态向量 $\boldsymbol{X}_t$ 都是 n 维的，随机过程向量 $\boldsymbol{U}_t$ 是 m 维的，并且 $\boldsymbol{A}_t$ 和 $\boldsymbol{B}_t$ 是和它们相对应的矩阵。

⊖ 向量的积分 $\int_0^t h(t,u)\mathrm{d}\boldsymbol{X}_u$ 可简单定义为随机向量，这个随机向量的第 j 个组成部分是 $\sum_{k=1}^{n}\int_0^t h_{jk}(t,u)\mathrm{d}X_u^{(k)}$，其中 h_{jk} 是 h 的第 jk 个元素并且 $X^{(k)}$ 是 $\boldsymbol{X}$ 的第 k 个元素。

含噪声的线性观测模型为

$$\boldsymbol{Y}_t = \boldsymbol{C}_t\boldsymbol{X}_t + \boldsymbol{N}_t,\quad 0 \leqslant t \leqslant T \tag{7.3.32b}$$

这里观测值 $\boldsymbol{Y}_t$ 和观测噪声 $\boldsymbol{N}_t$ 都是 r 维的，$\boldsymbol{C}_t$ 是 $t\times n$ 维的。

给定状态模型和观测模型后，估计状态值是关键，即在一定时间内($u\in[0,\ s]$)通过观测值 $\boldsymbol{Y}_u$ 估计出 $\boldsymbol{X}_t$，其中 $t<s$ 代表信号平滑，$t=s$ 代表信号滤波，$t>s$ 代表信号预测。与第 5 章以及本章前一部分描述类似，无论信号与噪声是否服从高斯假设，一旦明确其相关过程的二阶统计模型后，就可以得到最佳线性滤波，而且这些估计值是全局最优解。下面使用之前的方法进行推导。

假设状态驱动过程(也称为过程噪声)$\{\boldsymbol{U}_t;\ 0\leqslant t\leqslant T\}$和观测噪声$\{\boldsymbol{N}_t;\ 0\leqslant t\leqslant T\}$均为零均值不相关的单位白噪声，即

$$\mathrm{cov}(\boldsymbol{U}_t,\boldsymbol{U}_s) = \boldsymbol{I}_m\delta(t-s),\quad 0\leqslant t,s\leqslant T \tag{7.3.33a}$$

$$\mathrm{cov}(\boldsymbol{N}_t,\boldsymbol{N}_s) = \boldsymbol{I}_r\delta(t-s),\quad 0\leqslant t,s\leqslant T \tag{7.3.33b}$$

$$\mathrm{cov}(\boldsymbol{U}_t,\boldsymbol{N}_s) = \boldsymbol{0},\quad 0\leqslant t,s\leqslant T \tag{7.3.33c}$$

这里 $\boldsymbol{I}_m$ 和 $\boldsymbol{I}_r$ 分别是 $m\times m$ 和 $r\times r$ 阶单位矩阵，$\boldsymbol{0}$ 是全零矩阵。显然，对于所有满足 $t\neq s$ 的时刻 t、$s(t\neq s)$，噪声之间以及自身关于时间和组成元素都是不相关的；在 $t=s$ 时刻，它们之间以及自身关于组成元素不相关。同时假设矩阵函数$\{\boldsymbol{A}_t;\ 0\leqslant t\leqslant T\}$，$\{\boldsymbol{B}_t;\ 0\leqslant t\leqslant T\}$和$\{\boldsymbol{C}_t;\ 0\leqslant t\leqslant T\}$(分别称为状态矩阵、输入矩阵和观测矩阵)是关于时间的分段连续函数。

在给定初始状态下将式(7.3.31)改写为

$$\boldsymbol{X}_t = \boldsymbol{X}_0 + \int_0^t \boldsymbol{A}_u\boldsymbol{X}_u\,\mathrm{d}u + \int_0^t \boldsymbol{B}_u\,\mathrm{d}\boldsymbol{V}_u,\quad 0\leqslant t\leqslant T \tag{7.3.34}$$

过程 $\boldsymbol{V}_t \triangleq \int_0^t \boldsymbol{U}_s\,\mathrm{d}s$ 的自协方差为 $\boldsymbol{C}(t,\ u)=\boldsymbol{I}_m\min\{t,\ u\}$。对观测方程求积分得到

$$\boldsymbol{Z}_t \triangleq \int_0^t \boldsymbol{Y}_u\,\mathrm{d}u = \int_0^t \boldsymbol{C}_u\boldsymbol{X}_u\,\mathrm{d}u + \boldsymbol{W}_t\quad 0\leqslant t\leqslant T \tag{7.3.35}$$

其中 $\boldsymbol{W}_t$ 的协方差矩阵 $\boldsymbol{C}_w(t,\ u)=\boldsymbol{I}_r\min\{t,\ u\}$。在状态噪声和观测噪声都是白噪声时，模型 7.3.34 和模型 7.3.35 可简单表达为

$$\mathrm{d}\boldsymbol{X}_t = \boldsymbol{A}_t\boldsymbol{X}_t\,\mathrm{d}t + \boldsymbol{B}_t\,\mathrm{d}\boldsymbol{V}_t,\quad 0\leqslant t\leqslant T \tag{7.3.36}$$

和

$$\mathrm{d}\boldsymbol{Z}_t = \boldsymbol{C}_t\boldsymbol{X}_t\,\mathrm{d}t + \mathrm{d}\boldsymbol{W}_t,\quad 0\leqslant t\leqslant T \tag{7.3.37}$$

为了建立完整的统计模型，假定初始状态向量 $\boldsymbol{X}_0$ 是均值为零、协方差矩阵为 $\boldsymbol{P}_0$ 的随机向量，且与噪声过程 $\boldsymbol{V}_0^T$ 和 $\boldsymbol{W}_0^T$ 不相关。

在上述模型的基础上，进一步给出从观测值 $\boldsymbol{Z}_0^t$ 中推导 $\boldsymbol{X}_t$ 的递归方法。首先，观测方程(7.3.35)是式(7.3.23)在 $\boldsymbol{S}_u=C_u\boldsymbol{X}_u$ 时的一个特例。分别用$\hat{\boldsymbol{S}}_t$ 和$\hat{\boldsymbol{X}}_t$ 来表示 $\boldsymbol{S}_t$ 和 $\boldsymbol{X}_t$ 在观测值 $\boldsymbol{Z}_0^t$ 下的最佳线性估计，显然有$\hat{\boldsymbol{S}}_t=C_t\hat{\boldsymbol{X}}_t$。根据前面的讨论，可直接得到线性新息过程如下

$$\boldsymbol{I}_t = \boldsymbol{Z}_t - \int_0^t \boldsymbol{C}_u\hat{\boldsymbol{X}}_u\,\mathrm{d}u,\quad 0\leqslant t\leqslant T \tag{7.3.38}$$

该新息过程与$\{\boldsymbol{W}_t;\ 0\leqslant t\leqslant T\}$一样，它的增量 $\boldsymbol{I}_t-\boldsymbol{I}_s$ 与过去的观测值 $\boldsymbol{Z}_0^s$ 不相关。更重要的是，该线性新息过程和用来进行线性估计的观测值等价，所以可将 $\boldsymbol{X}_t$ 写成如下形式

$$\hat{\boldsymbol{X}}_t = \int_0^t \boldsymbol{g}(t,u)\mathrm{d}\boldsymbol{I}_u \tag{7.3.39}$$

其中 $\boldsymbol{g}$ 是矩阵形式的冲激响应。

求解 $\boldsymbol{g}$ 需要利用正交性原理，即

$$E\{(\boldsymbol{X}_t - \hat{\boldsymbol{X}}_t)\boldsymbol{I}_u{}'\} = 0,\quad 0 \leqslant u \leqslant t \tag{7.3.40}$$

合并式(7.3.39)和式(7.3.40)有

$$E\{\boldsymbol{X}_t\boldsymbol{I}'_u\} = E\left\{\left(\int_0^t \boldsymbol{g}(t,v)\mathrm{d}\boldsymbol{I}_v\right)\boldsymbol{I}'_u\right\}$$

该式等价为

$$\boldsymbol{C}_{XI}(t,u) = \int_0^u \boldsymbol{g}(t,v)\mathrm{d}v,\quad 0 \leqslant u \leqslant t$$

对上式求微分可得

$$\boldsymbol{g}(t,u) = \frac{\partial}{\partial u}\boldsymbol{C}_{XI}(t,u),\quad 0 \leqslant u \leqslant t \tag{7.3.41}$$

需要指出的是，式(7.3.41)是不依赖状态模型(7.3.34)的通解。注意

$$\boldsymbol{I}_u = \int_0^u \boldsymbol{C}_v(\boldsymbol{X}_v - \hat{\boldsymbol{X}}_v)\mathrm{d}v + \boldsymbol{W}_u \tag{7.3.42}$$

可将 $\boldsymbol{C}_{XI}(t,u)$ 表达为

$$\boldsymbol{C}_{XI}(t,u) \triangleq E\{\boldsymbol{X}_t\boldsymbol{I}'_u\} = \int_0^u E\{\boldsymbol{X}_t(\boldsymbol{X}_v - \hat{\boldsymbol{X}}_v)'\}\boldsymbol{C}'_v\mathrm{d}v + E\{\boldsymbol{X}_t\boldsymbol{W}'_u\} \tag{7.3.43}$$

因为 $\boldsymbol{X}_t$ 由 $\boldsymbol{X}_0$ 和 $\boldsymbol{V}_0^t$ 线性表出，而这两者与 W_0^T 都不相关，因此有 $E\{\boldsymbol{X}_t\boldsymbol{W}'_u\}=0$。这样，结合式(7.3.41)和式(7.3.43)可得到

$$\boldsymbol{g}(t,u) = E\{\boldsymbol{X}_t(\boldsymbol{X}_u - \hat{\boldsymbol{X}}_u)'\}\boldsymbol{C}'_u,\quad 0 \leqslant u \leqslant t \leqslant T \tag{7.3.44}$$

根据式(7.3.44)可以得到求解 $\hat{\boldsymbol{X}}_t$ 的递归计算式。在给出具体的递归形式之前，首先将 $\boldsymbol{g}$ 进行如下变换

$$\boldsymbol{g}(t,u) = E\{(\boldsymbol{X}_t - \boldsymbol{X}_u)(\boldsymbol{X}_u - \hat{\boldsymbol{X}}_u)'\}\boldsymbol{C}'_u + E\{\boldsymbol{X}_u(\boldsymbol{X}_u - \hat{\boldsymbol{X}}_u)'\}\boldsymbol{C}'_u$$

根据状态方程(7.3.34)有

$$\boldsymbol{X}_t - \boldsymbol{X}_u = \int_u^t \boldsymbol{A}_v\boldsymbol{X}_v\mathrm{d}v + \int_u^t \boldsymbol{B}_v\mathrm{d}\boldsymbol{V}_v$$

进一步可以得到

$$\boldsymbol{g}(t,u) = \int_u^t \boldsymbol{A}_v E\{\boldsymbol{X}_v(\boldsymbol{X}_u - \hat{\boldsymbol{X}}_u)'\}\boldsymbol{C}'_u\mathrm{d}v + E\left\{\left(\int_u^t \boldsymbol{B}_v\mathrm{d}\boldsymbol{V}_v\right)(\boldsymbol{X}_u - \hat{\boldsymbol{X}}_u)'\right\} + E\{\boldsymbol{X}_u(\boldsymbol{X}_u - \hat{\boldsymbol{X}}_u)'\}\boldsymbol{C}'_u \tag{7.3.45}$$

根据式(7.3.44)，可知式(7.3.45)右边第一部分就是 $\int_u^t \boldsymbol{A}_v\boldsymbol{g}(v,u)\mathrm{d}v$ 。一方面，因为 $\boldsymbol{X}_u$ 和 $\hat{\boldsymbol{X}}_u$ 均由与增量 $\{\boldsymbol{V}_v - \boldsymbol{V}_u;\ u \leqslant v \leqslant T\}$ 无关的 $\boldsymbol{X}_0$、$\boldsymbol{V}_0^u$ 和 $\boldsymbol{W}_0^u$ 线性构成；另一方面，$\int_u^t \boldsymbol{B}_v\mathrm{d}\boldsymbol{V}_v$ 是增量的线性表示，所以上式的第二部分等于零。第三部分仅仅和 u 有关，记为 K_u。这样 $\boldsymbol{g}(t,u)$ 可表达为

$$\boldsymbol{g}(t,u) = \int_u^t \boldsymbol{A}_v\boldsymbol{g}(v,u)\mathrm{d}v + \boldsymbol{K}_u,\quad 0 \leqslant u \leqslant t \leqslant T \tag{7.3.46}$$

进而可得到

$$\hat{\boldsymbol{X}}_t=\int_0^t \boldsymbol{g}(t,u)\mathrm{d}\boldsymbol{I}_u=\int_0^t\int_u^t \boldsymbol{A}_v\boldsymbol{g}(v,u)\mathrm{d}v\mathrm{d}\boldsymbol{I}_u+\int_0^t \boldsymbol{K}_u\mathrm{d}\boldsymbol{I}_u=\int_0^t \boldsymbol{A}_v\int_0^v \boldsymbol{g}(v,u)\mathrm{d}\boldsymbol{I}_u\mathrm{d}v+\int_0^t \boldsymbol{K}_u\mathrm{d}\boldsymbol{I}_u$$
$$=\int_0^t \boldsymbol{A}_v\hat{\boldsymbol{X}}_v\mathrm{d}v+\int_0^t \boldsymbol{K}_u\mathrm{d}\boldsymbol{I}_u \tag{7.3.47}$$

上式可简单表示为

$$\mathrm{d}\hat{\boldsymbol{X}}_t=\boldsymbol{A}_t\hat{\boldsymbol{X}}_t+\boldsymbol{K}_t\mathrm{d}\boldsymbol{I}_t,\quad 0\leqslant t\leqslant T \tag{7.3.48}$$

其中，初始条件为$\hat{\boldsymbol{X}}_0=\boldsymbol{0}$。

由式(7.3.48)给出的最优状态估计器与式(7.3.36)的动态特性相同，即具有相同的状态矩阵，但输入矩阵添加了$\boldsymbol{K}_t$。该方程由统计特性与白噪声积分后相似的新息过程驱动。因此，描述状态估计器的方程和原始的状态方程相似。类似地，在离散信号形式下也有相同的结论。

根据式(7.3.48)求解状态估计量$\hat{\boldsymbol{X}}_t$，需确定增益矩阵$\boldsymbol{K}_t$的表达式。根据定义，有$\boldsymbol{K}_t=E\{\boldsymbol{X}_t(\boldsymbol{X}_t-\hat{\boldsymbol{X}}_t)\}\boldsymbol{C}'_t$，由于$\hat{\boldsymbol{X}}_t$可由$\boldsymbol{I}_0^t$线性表示，所以它和估计误差$(\boldsymbol{X}_t-\hat{\boldsymbol{X}}_t)$是不相关的。因此，$\boldsymbol{K}_t$可改写为

$$\boldsymbol{K}_t=E\{(\hat{\boldsymbol{X}}_t-\boldsymbol{X}_t)(\hat{\boldsymbol{X}}_t-\boldsymbol{X}_t)'\}\boldsymbol{C}'_t=\boldsymbol{P}_t\boldsymbol{C}'_t \tag{7.3.49}$$

其中

$$\boldsymbol{P}_t\triangleq E\{(\boldsymbol{X}_t-\hat{\boldsymbol{X}}_t)(\boldsymbol{X}_t-\hat{\boldsymbol{X}}_t)'\}$$

上式是估计误差在t时刻的协方差矩阵。因此，通过确定$\boldsymbol{P}_t$便可以得到增益$\boldsymbol{K}_t$的具体表达式。

设误差过程$\boldsymbol{e}_t\triangleq\boldsymbol{X}_t-\hat{\boldsymbol{X}}_t$，其自协方差函数用$\boldsymbol{C}_e(t,u)$来表示。根据式(7.3.34)和式(7.3.47)，$\{\boldsymbol{e}_t;\ t\in[0,T]\}$满足

$$\boldsymbol{e}_t=\boldsymbol{e}_0+\int_0^t \boldsymbol{A}_u\boldsymbol{e}_u\mathrm{d}u+\int_0^t \boldsymbol{B}_u\mathrm{d}\boldsymbol{V}_u-\int_0^t \boldsymbol{K}_u\mathrm{d}\boldsymbol{I}_u,\quad 0\leqslant t\leqslant T \tag{7.3.50}$$

其中$\boldsymbol{e}_0=\boldsymbol{X}_0-\hat{\boldsymbol{X}}_0=\boldsymbol{X}_0$。在该式基础上可以推出微分方程，进而可以得到$\boldsymbol{P}_t\triangleq\boldsymbol{C}_e(t,t)$。为便于分析，首先假设$\boldsymbol{e}_0\equiv\boldsymbol{0}$（或者假设$P_0=0$，两者等价）。在该假设下，结合式(7.3.50)，有

$$\boldsymbol{P}_t\triangleq E\{\boldsymbol{e}_t\boldsymbol{e}'_t\}=\int_0^t \boldsymbol{A}_u E\{\boldsymbol{e}_u\boldsymbol{e}'_t\}\mathrm{d}u+E\left\{\left(\int_0^t \boldsymbol{B}_u\mathrm{d}\boldsymbol{V}_u-\int_0^t \boldsymbol{K}_u\mathrm{d}\boldsymbol{I}_u\right)\boldsymbol{e}'_t\right\}$$
$$=\int_0^t \boldsymbol{A}_u\boldsymbol{C}_e(u,t)\mathrm{d}u+E\left\{\left(\int_0^t \boldsymbol{B}_u\mathrm{d}\boldsymbol{V}_u\right)\boldsymbol{e}'_t\right\} \tag{7.3.51}$$

其中，利用了在所有$u\in[0,t]$下$\boldsymbol{e}_t$和$\boldsymbol{I}_u$正交的性质。再次利用式(7.3.50)后，可以进一步得到

$$\boldsymbol{P}_t=\int_0^t\int_0^t \boldsymbol{A}_u\boldsymbol{C}_e(u,v)\boldsymbol{A}'_v\mathrm{d}u\mathrm{d}v+\int_0^t \boldsymbol{A}_u E\left\{\boldsymbol{e}_u\left(\int_0^t \boldsymbol{B}_v\mathrm{d}\boldsymbol{V}_v-\int_0^t \boldsymbol{K}_v\mathrm{d}\boldsymbol{I}_v\right)'\right\}\mathrm{d}u$$
$$+E\left\{\int_0^t \boldsymbol{B}_u\mathrm{d}\boldsymbol{V}_u\left(\int_0^t \boldsymbol{A}_v\boldsymbol{e}_v\mathrm{d}v\right)'\right\}+E\left\{\int_0^t \boldsymbol{B}_u\mathrm{d}\boldsymbol{V}_u\left(\int_0^t \boldsymbol{B}_v\mathrm{d}\boldsymbol{V}_v-\int_0^t \boldsymbol{K}_v\mathrm{d}\boldsymbol{I}_v\right)'\right\}$$
$$=\int_0^t\int_0^t \boldsymbol{A}_u\boldsymbol{C}_e(u,v)\boldsymbol{A}'_v\mathrm{d}u\mathrm{d}v+\int_0^t \boldsymbol{A}_u E\left\{\boldsymbol{e}_u\left(\int_0^t \boldsymbol{B}_v\mathrm{d}\boldsymbol{V}_v\right)'\right\}\mathrm{d}u$$
$$+\int_0^t E\left\{\left(\int_0^t \boldsymbol{B}_u\mathrm{d}\boldsymbol{V}_u\right)\boldsymbol{e}'_v\right\}\boldsymbol{A}'_v\mathrm{d}v+E\left\{\left(\int_0^t \boldsymbol{B}_u\mathrm{d}\boldsymbol{V}_u\right)\left(\int_0^t \boldsymbol{B}_v\mathrm{d}\boldsymbol{V}_v\right)'\right\}$$

$$-E\left\{\left(\int_0^t \boldsymbol{A}_u\boldsymbol{e}_u\mathrm{d}u+\int_0^t \boldsymbol{B}_u\mathrm{d}\boldsymbol{V}_u\right)\left(\int_0^t \boldsymbol{K}_v\mathrm{d}\boldsymbol{I}_v\right)'\right\} \tag{7.3.52}$$

上式中，第二个等式是将第一个等式中包含 $\int_0^t \boldsymbol{K}_v\mathrm{d}\boldsymbol{I}_v$ 的两项合并得到。

考虑上式的最后一项，利用式(7.3.50)可得

$$E\left\{\left(\int_0^t \boldsymbol{A}_u\boldsymbol{e}_u\mathrm{d}u+\int_0^t \boldsymbol{B}_u\mathrm{d}\boldsymbol{V}_u\right)\left(\int_0^t \boldsymbol{K}_v\mathrm{d}\boldsymbol{I}_v\right)'\right\}=E\left\{\left(\boldsymbol{e}_t+\int_0^t \boldsymbol{K}_u\mathrm{d}\boldsymbol{I}_u\right)\left(\int_0^t \boldsymbol{K}_v\mathrm{d}\boldsymbol{I}_v\right)'\right\}$$
$$=E\left\{\left(\int_0^t \boldsymbol{K}_u\mathrm{d}\boldsymbol{I}_u\right)\left(\int_0^t \boldsymbol{K}_v\mathrm{d}\boldsymbol{I}_v\right)'\right\}=\int_0^t \boldsymbol{K}_u\boldsymbol{K}'_u\mathrm{d}u \tag{7.3.53}$$

这里第二个等式利用了 $\boldsymbol{e}_t$ 和 $\boldsymbol{I}_0^t$ 的正交性，第三个等式利用了 $\boldsymbol{C}_I(u,\ v)=\boldsymbol{I}_r\min\{u,\ v\}$。而且，因为 $\boldsymbol{e}_u$ 由 $\boldsymbol{e}_0$、$\boldsymbol{V}_0^u$ 和 $\boldsymbol{W}_0^u$ 获得，所以它在所有 $s\geqslant u$ 下和 $(\boldsymbol{V}_s-\boldsymbol{V}_u)$ 正交。这样式(7.3.52)右边的第二项可写为

$$\int_0^t \boldsymbol{A}_u E\left\{\boldsymbol{e}_u\left(\int_0^t \boldsymbol{B}_v\mathrm{d}\boldsymbol{V}_v\right)'\right\}\mathrm{d}u=\int_0^t \boldsymbol{A}_u E\left\{\boldsymbol{e}_u\left(\int_0^u \boldsymbol{B}_v\mathrm{d}\boldsymbol{V}_v\right)'\right\}\mathrm{d}u \tag{7.3.54}$$

类似地，第三项也成立。将式(7.3.53)和式(7.3.54)代入式(7.3.52)中，并利用 $\boldsymbol{C}_V(u,\ v)=\boldsymbol{I}_m\min\{u,\ v\}$后，有

$$\boldsymbol{P}_t=\int_0^t\int_0^t \boldsymbol{A}_u\boldsymbol{C}_e(u,v)\boldsymbol{A}'_v\mathrm{d}u\mathrm{d}v+\int_0^t \boldsymbol{A}_u E\left\{\boldsymbol{e}_u\left(\int_0^u \boldsymbol{B}_v\mathrm{d}\boldsymbol{V}_v\right)'\right\}\mathrm{d}u$$
$$+\int_0^t E\left\{\left(\int_0^v \boldsymbol{B}_u\mathrm{d}\boldsymbol{V}_u\right)\boldsymbol{e}'_v\right\}\boldsymbol{A}'_v\mathrm{d}v+\int_0^t \boldsymbol{B}_u\boldsymbol{B}'_u\mathrm{d}u-\int_0^t \boldsymbol{K}_u\boldsymbol{K}'_u\mathrm{d}u,\quad 0\leqslant t\leqslant T \tag{7.3.55}$$

对式(7.3.55)两边进行微分，可得

$$\dot{\boldsymbol{P}}_t=\int_0^t \boldsymbol{A}_t\boldsymbol{C}_e(t,v)\boldsymbol{A}'_v\mathrm{d}v$$
$$+\int_0^t \boldsymbol{A}_u\boldsymbol{C}_e(u,t)\boldsymbol{A}'_t\mathrm{d}u+\boldsymbol{A}_t E\left\{\boldsymbol{e}_t\left(\int_0^t \boldsymbol{B}_v\mathrm{d}\boldsymbol{V}_v\right)'\right\}+E\left\{\left(\int_0^t \boldsymbol{B}_u\mathrm{d}\boldsymbol{V}_u\right)\boldsymbol{e}'_t\right\}\boldsymbol{A}'_t+\boldsymbol{B}_t\boldsymbol{B}'_t-\boldsymbol{K}_t\boldsymbol{K}'_t$$
$$=\boldsymbol{A}_t\left[\int_0^t \boldsymbol{C}_e(t,v)\boldsymbol{A}'_v\mathrm{d}v+E\left\{\boldsymbol{e}_t\left(\int_0^t \boldsymbol{B}_v\mathrm{d}\boldsymbol{V}_v\right)'\right\}\right]$$
$$+\left[\int_0^t \boldsymbol{A}_u\boldsymbol{C}_e(u,t)\mathrm{d}u+E\left\{\left(\int_0^t \boldsymbol{B}_u\mathrm{d}\boldsymbol{V}_u\right)\boldsymbol{e}'_t\right\}\right]\boldsymbol{A}'_t+\boldsymbol{B}_t\boldsymbol{B}'_t-\boldsymbol{K}_t\boldsymbol{K}'_t$$
$$=\boldsymbol{A}_t\boldsymbol{P}'_t+\boldsymbol{P}_t\boldsymbol{A}'_t+\boldsymbol{B}_t\boldsymbol{B}'_t-\boldsymbol{K}_t\boldsymbol{K}'_t \tag{7.3.56}$$

第三个等式由式(7.3.51)得到。根据 $\boldsymbol{P}'_t=\boldsymbol{P}_t$ 和 $\boldsymbol{K}_t=\boldsymbol{P}_t\boldsymbol{C}'_t$，式(7.3.56)可改写为

$$\dot{\boldsymbol{P}}_t=\boldsymbol{A}_t\boldsymbol{P}'_t+\boldsymbol{P}_t\boldsymbol{A}'_t+\boldsymbol{B}_t\boldsymbol{B}'_t-\boldsymbol{P}_t\boldsymbol{C}'_t\boldsymbol{C}_t\boldsymbol{P}_t,\quad 0\leqslant t\leqslant T \tag{7.3.57}$$

上式是 $\boldsymbol{P}_t$ 的微分方程。根据之前的假设，该方程的初始条件是 $\boldsymbol{P}_0=0$。但是，对任意的初始状态协方差矩阵 $\boldsymbol{P}_0$ 而言，式(7.3.57)同样适用，具体推导过程留作练习。

在总结连续时域 Kalman-Bucy 滤波器之前，先对两个假设条件进行部分修改。第一个是将条件式(7.3.33a)和式(7.3.33b)放宽为

$$\mathrm{cov}(\boldsymbol{U}_t,\boldsymbol{U}_s)=\boldsymbol{Q}_t\delta(t-s),\quad 0\leqslant t,s\leqslant T \tag{7.3.58a}$$

和

$$\mathrm{cov}(\boldsymbol{N}_t,\boldsymbol{N}_s)=\boldsymbol{R}_t\delta(t-s),\quad 0\leqslant t,s\leqslant T \tag{7.3.58b}$$

对于每个 $t\in[0,\ T]$，$\boldsymbol{Q}_t$ 和 $\boldsymbol{R}_t$ 表示对应时刻的协方差矩阵且 $\boldsymbol{R}_t>0$。在假设这些矩阵函数

都是关于 t 的分段连续函数之后，只需对滤波器形式进行简单修改便能满足上式。特别地，类似第 3 章中的分析，可将 $\boldsymbol{Q}_t$ 和 $\boldsymbol{R}_t$ 写成

$$\boldsymbol{Q}_t = \boldsymbol{Q}_t^{1/2}\boldsymbol{Q}_t^{1/2}, \quad \boldsymbol{R}_t = \boldsymbol{R}_t^{1/2}\boldsymbol{R}_t^{1/2}$$

其中，$\boldsymbol{Q}_t^{1/2}$ 和 $\boldsymbol{R}_t^{1/2}$ 是对称矩阵且 $\boldsymbol{R}_t^{1/2}>0$。进而可将 $\boldsymbol{U}_t$ 表示为由协方差为 $\boldsymbol{I}_m\delta(t-s)$ 的白噪声乘以 $\boldsymbol{Q}_t^{1/2}$ 的形式，$\boldsymbol{N}_t$ 是协方差为 $\boldsymbol{I}_r\delta(t-s)$ 的白噪声的 $\boldsymbol{R}_t^{1/2}$ 倍。又因为 $\boldsymbol{R}_t^{1/2}$ 可逆，所以在滤波器方程中只需用 $\boldsymbol{B}_t\boldsymbol{Q}_t^{1/2}$ 替代 $\boldsymbol{B}_t$、$\boldsymbol{P}_t\boldsymbol{C}_t^T\boldsymbol{R}_t^{-1}$ 替代 $\boldsymbol{K}_t$、$\boldsymbol{R}_t^{1/2}\boldsymbol{C}_t$ 替代 $\boldsymbol{C}_t$。

第二个关于假设的修正是将 $E\{\boldsymbol{X}_0\}=0$ 改写为 $E\{\boldsymbol{X}_0\}=\boldsymbol{m}_0$。该修正仅改变了随机过程 $\{\boldsymbol{X}_t;\ 0\leqslant t\leqslant T\}$ 和 $\{\boldsymbol{Y};\ 0\leqslant t\leqslant T\}$ 的均值，与第 5 章中讨论的情况类似，均值非零情况下的 MMSE 线性估计只需对零均值情况进行简单修正即可。在 Kalman-Bucy 模型中，非零均值假设只需将估计方程的初始条件改成 $\hat{\boldsymbol{X}}_0=\boldsymbol{m}_0$。

作为对上述内容的总结，以命题的形式给出连续时间域的 Kalman-Bucy 模型及其最优估计量。

命题 7.3.2 连续时间 Kalman-Bucy 滤波器的解

假设 $\{\boldsymbol{X}_t;\ 0\leqslant t\leqslant T\}$ 和 $\{\boldsymbol{Z}_t;\ 0\leqslant t\leqslant T\}$ 由以下模型产生

$$\mathrm{d}\boldsymbol{X}_t = \boldsymbol{A}_t\boldsymbol{X}_t\mathrm{d}t + \boldsymbol{B}_t\mathrm{d}\boldsymbol{V}_t, \quad 0\leqslant t\leqslant T \tag{7.3.59a}$$

$$\mathrm{d}\boldsymbol{Z}_t = \boldsymbol{C}_t\boldsymbol{X}_t\mathrm{d}t + \mathrm{d}\boldsymbol{W}_t, \quad 0\leqslant t\leqslant T \tag{7.3.59b}$$

其中，初始条件为 $\boldsymbol{X}_0$，对于所有 $0\leqslant t,\ s\leqslant T$，有 $E\{\boldsymbol{V}_t\}=E\{\boldsymbol{W}_t\}=\boldsymbol{0}$，$\boldsymbol{C}_V(t,s)=\int_0^{\min\{t,s\}}\boldsymbol{Q}_u\mathrm{d}u$，$\boldsymbol{C}_W(t,s)=\int_0^{\min\{t,s\}}\boldsymbol{R}_u\mathrm{d}u$，$\boldsymbol{C}_{VW}(t,\ s)=0$，$\boldsymbol{A}_t$、$\boldsymbol{B}_t$、$\boldsymbol{C}_t$、$\boldsymbol{Q}_t$ 和 $\boldsymbol{R}_t$ 都是关于 t 的分段连续函数，且在 $t\in[0,\ T]$ 时 $\boldsymbol{R}_t>0$。进一步假设初始条件 $\boldsymbol{X}_0$ 的均值为 $\boldsymbol{m}_0$、协方差为 $\boldsymbol{P}_0$，且与 $\boldsymbol{V}_0^T$、$\boldsymbol{W}_0^T$ 不相关。在给定观测数据 $\boldsymbol{Z}_0^t$ 下 $\boldsymbol{X}_t$ 的 MMSE 线性估计为

$$\mathrm{d}\hat{\boldsymbol{X}}_t = \boldsymbol{A}_t\hat{\boldsymbol{X}}_t\mathrm{d}t + \boldsymbol{K}_t\mathrm{d}\boldsymbol{I}_t, \quad \hat{\boldsymbol{X}}_0=\boldsymbol{m}_0 \tag{7.3.60}$$

其中，$\mathrm{d}\boldsymbol{I}_t=\mathrm{d}\boldsymbol{Z}_t-\boldsymbol{C}_t\hat{\boldsymbol{X}}_t\mathrm{d}t$，且有

$$\boldsymbol{K}_t = \boldsymbol{P}_t\boldsymbol{C}'_t\boldsymbol{R}_t^{-1} \tag{7.3.61}$$

$\boldsymbol{P}_t\triangleq\boldsymbol{E}\{(\boldsymbol{X}_t-\hat{\boldsymbol{X}}_t)(\boldsymbol{X}_t-\hat{\boldsymbol{X}}_t)'\}$ 满足 Riccati 微分方程，即

$$\dot{\boldsymbol{P}}_t = \boldsymbol{A}_t\boldsymbol{P}_t + \boldsymbol{P}_t\boldsymbol{A}'_t + \boldsymbol{B}_t\boldsymbol{Q}_t\boldsymbol{B}'_t - \boldsymbol{P}_t\boldsymbol{C}'_t\boldsymbol{R}_t^{-1}\boldsymbol{C}_t\boldsymbol{P}_t \tag{7.3.62}$$

初始条件为 $\boldsymbol{P}_0$。

和离散情况相似，该最佳滤波器可由新息过程经 $\boldsymbol{K}_t$ 倍增益后驱动原始系统得到，或者由观测值 $\mathrm{d}\boldsymbol{Z}_t$ 经 $\boldsymbol{K}_t$ 倍增益后驱动状态矩阵为 $\{(\boldsymbol{A}_t-\boldsymbol{K}_t\boldsymbol{C}_t),\ 0\leqslant t\leqslant T\}$ 的系统得到。将 $\mathrm{d}\boldsymbol{Z}_t$ 等价为 $\boldsymbol{Y}_t\mathrm{d}t$ 后（$\boldsymbol{Y}_t$ 是白噪声污染下的观测值），图 7.3.1 给出了 Kalman-Bucy 滤波器的具体实现结构。在习题中涉及了更多关于连续时间域 Kalman-Bucy 滤波的问题，更详细的讨论可以参考 Van Trees(1968) 和 Kailath(1981) 的著述。

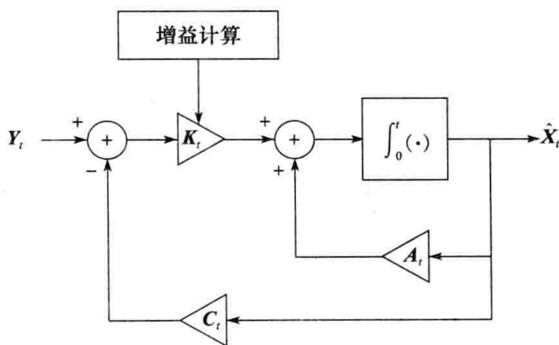


图 7.3.1 连续时间域 Kalman-Bucy 滤波器

7.3.4　线性/高斯问题的进一步扩展

与 5.3 节讨论的离散情形一样，线性/高斯估计可适用于更加一般的情况。为便于讨论，假设两个二阶随机过程$\{X_t;\ t\in\mathbf{R}\}$和$\{Y_t;\ t\in\mathbf{R}\}$的均值为零，并通过在时间间隔$[a,\ b]$内的某些观测值$\{Y_t;\ t\in\mathbf{R}\}$来估计X_t。

首先，记X_t的所有线性估计构成的集合H_a^b为

$$\sum_{i=1}^{n} c_i Y_{t_i}$$

其中，$c_i\in\mathbf{R}$，$t_i\in[a,\ b]$，$i=1,\ \cdots,\ n$ 且均方可和。与之对应的线性 MMSE 估计可表达为

$$\min_{Z\in H_a^b} E\{(Z-X_t)^2\} \tag{7.3.63}$$

正交性是$\hat{X}$为最佳估计值的充要条件，即

$$E\{(X_t-\hat{X}_t)Z\}=0,\quad Z\in H_a^b \tag{7.3.64}$$

该式等价于

$$E\{(\hat{X}_t-X_t)Y_s\}=0,\quad a\leqslant s\leqslant b \tag{7.3.65}$$

对$\{Y_t;\ a\leqslant t\leqslant b\}$线性滤波后可构成$H_a^b$的一类子集，即

$$\int_a^b h(t,\tau)Y_\tau \mathrm{d}\tau \tag{7.3.66}$$

其中，$\int_a^b\int_a^b h(t,\tau)h(t,\sigma)C_Y(\tau,\sigma)\mathrm{d}\tau\mathrm{d}\sigma<\infty$（$C_Y$ 表示$\{Y_t;\ t\in\mathbf{R}\}$的自协方差函数）。利用式(7.3.65)并交换均方极限与期望的顺序，可知当且仅当 h 满足如下积分方程时，由式(7.3.66)给出的线性估计量是最优的

$$C_{XY}(t,s)=\int_a^b h(t,\tau)C_Y(\tau,s)\mathrm{d}\tau,\quad a\leqslant s\leqslant b \tag{7.3.67}$$

其中，C_{XY}是$\{X_t;\ t\in\mathbf{R}\}$和$\{Y_t;\ t\in\mathbf{R}\}$的互协方差函数。

方程(7.3.67)是式(5.3.18)中 Wiener-Hopf 方程在连续时间域的表达。该方程是第一类 Fredholm 方程，它的解具有奇异性。与离散域不同，除非在$h(t,\ \tau)$中允许存在奇异点，否则式(7.3.66)仅能对部分线性 MMSE 估计问题建模。例如，估计值可能不仅包含$\{Y_t;\ a\leqslant t\leqslant b\}$的滤波形式，还包含对$\{Y_t;\ a\leqslant t\leqslant b\}$以及其导数的样本加权。

对于式(7.3.66)给出的估计量是否可以包含大多数感兴趣的解，需要分两种情况进行讨论，第一种是观测间隔无穷大(也就是$a=-\infty$或者$b=+\infty$)，第二种是$\{Y_t;\ t\in\mathbf{R}\}$中含有与$\{X_t;\ t\in\mathbf{R}\}$不相关的白噪声。

无限观测问题中最常见的是连续时间域的 Wiener-Kolmogorov 问题。分别假设$\{X_t;\ t\in\mathbf{R}\}$和$\{Y_t;\ t\in\mathbf{R}\}$是独立的联合广义平稳随机变量(即 C_{XY}和 C_Y 仅是参数的差的函数，仅与时间间隔有关，而与时间起点无关)，在$a=-\infty$的条件下，$b=t$ 和$b=+\infty$分别对应问题的因果和非因果两种情形。不管在哪一种情况下，都可以考虑为时不变滤波器的问题。例如，对于某些函数$h:\mathbf{R}\to\mathbf{R}$，$h(t,\ \tau)$可以写成$h(t-\tau)$的形式。在滤波器满足时不变的条件下，非因果的 Wiener-Hopf 方程可简化为

$$C_{XY}(\tau)=\int_{-\infty}^{\infty}h(\alpha)C_Y(\tau-\alpha)\mathrm{d}\alpha,\quad -\infty<\tau<\infty \tag{7.3.68}$$

因果方程可简化为

$$C_{XY}(\tau)=\int_{0}^{\infty}h(\alpha)C_Y(\tau-\alpha)\mathrm{d}\alpha,\quad 0\leqslant\tau<\infty \tag{7.3.69}$$

和离散情况类似，式(7.3.68)和式(7.3.69)都可以通过傅里叶变换来求解。假设 C_{XY}、h 和 C_Y 的傅里叶变换分别是 ϕ_{XY}、H 和 ϕ_Y(注意，ϕ_{XY}是$\{X_t;\ t\in\mathbf{R}\}$和$\{Y_t;\ t\in\mathbf{R}\}$的互功率谱，ϕ_Y 是$\{Y_t;\ t\in\mathbf{R}\}$的功率谱)，考虑到式(7.3.68)可表达为 $C_{XY}(\tau)=(h*C_Y)(\tau)$(其中 $*$ 代表卷积运算)，则式(7.3.68)等价为

$$\phi_{XY}(\omega)=H(\omega)\phi_Y(\omega),\quad -\infty<\omega<\infty \tag{7.3.70}$$

显然，最佳非因果滤波器的传递函数为

$$H_0(\omega)=\frac{\phi_{XY}(\omega)}{\phi_Y(\omega)} \tag{7.3.71}$$

在对滤波响应施加因果约束后，求解式(7.3.69)存在一定的难度。类似于离散情形，需要将观测谱 ϕ_Y 因式分解成因果部分 ϕ_Y^+ 和非因果部分 $\phi_Y^-[\equiv(\phi_Y^+)^*]$的形式。可以证明 ϕ_Y 满足连续时间域 Paley-Wiener 条件[参考 Wang(1983)]，即

$$\int_{-\infty}^{\infty}|\log\phi_Y(\omega)|/(1+\omega^2)\mathrm{d}\omega<\infty \tag{7.3.72}$$

对 ϕ_Y 进行因式分解后，式(7.3.70)的传递函数可表示为

$$H_0^+(\omega)=\frac{1}{\phi_Y^+(\omega)}\left[\frac{\phi_{XY}(\omega)}{\phi_Y^-(\omega)}\right]_+ \tag{7.3.73}$$

这里运算$[\cdot]_+$表示取谱的因果部分，即$[H(\omega)]_+=F\{h^+(t)\}$，其中

$$h^+(t)=\begin{cases}h(t), & t\geqslant 0\\ 0, & t<0\end{cases} \tag{7.3.74}$$

将式(7.3.73)中各部分表达式分别代入后即可证明式(7.3.73)是式(7.3.69)的解。显然，同离散时间滤波器的直观解释一致，该滤波器可看作由因果预白化滤波器 $1/\Phi_Y^+$ 和因果 MMSE 估计器级联而成，其中因果 MMSE 估计器用来从预白化后的观测数据中估计 X_t。

第二类估计问题是观测量$\{Y_t;\ t\in\mathbf{R}\}$中包含与$\{X_t;\ t\in\mathbf{R}\}$不相关的白噪声，它可根据式(7.3.67)描述的 Wiener-Hopf 方程直接求解。为不失一般性，假设 $C_Y(t,s)=C_c(t,s)+(N_0/2)\delta(t-s)$，则 Wiener-Hopf 方程可写为

$$C_{XY}(t,s)=\int_a^b h(t,\tau)C_c(\tau,s)\mathrm{d}\tau+\frac{N_0}{2}h(t,s),\quad a\leqslant s\leqslant b \tag{7.3.75}$$

上式是第二类 Fredholm 方程，当 C_{XY}和 C_c 连续且 $N_0/2>0$ 时存在唯一解。

下面以非相关加性白噪声中估计信号为例来说明 Wiener-Hopf 方程在线性/高斯估计问题中的重要性。设 $X_t=S_t$ 以及

$$X_t=S_t+N_t,\quad a\leqslant t\leqslant b \tag{7.3.76}$$

其中，$\{N_t;\ -\infty<t<\infty\}$为白噪声信号，$C_{XY}=C_c=C_x$。显然，与之对应的 Wiener-Hopf 方程可简化为式(7.3.7)。事实上，Kalman-Bucy 滤波问题是式(7.3.76)的特例，且直接

从 Wiener-Hopf 方程就能得到 Kalman-Bucy 滤波器[参考 Van Trees(1968)]。这是因为方程(7.3.41)在本质上就是利用新息过程来估计状态变量的 Wiener-Hopf 方程。

7.4 非线性滤波

对于第7.3节中讨论的线性/高斯信号估计问题已经有了相对较好的解决方式。然而，当线性约束条件和高斯模型不再成立时，连续时间域的最优信号估值问题变得相当困难。此时，信号估计所面临的基本问题与连续时间域的非高斯信号检测问题一致，也就是如何在函数空间中构造适合于非高斯测量的概率密度函数。不过，在对这个问题附加某些约束后，特别是假设信号和观测均由非线性动态模型产生(例如 Kalman-Bucy 滤波中所涉及的非线性条件)，就能够很好地近似大部分的非高斯物理现象。因此，本节将重点考虑非线性动态模型的最优滤波问题。对非线性问题进行完整和严格的阐述是非常复杂的，完全可以单独成书，所以本节仅简述其核心思想。读者若是对具体细节问题感兴趣，可以参考 Lipster 和 Shiryayev 的书(1977)，以及 Bebes(1987)和 Marcus(1984)写的综述。

大量物理过程的演化特性可以建模为非线性随机微分方程的形式，即

$$\mathrm{d}X_t = m(X_t,t)\mathrm{d}t + \sigma(X_t,t)\mathrm{d}W_t,\quad t \geqslant 0 \tag{7.4.1}$$

其中，状态的初始值为 X_0，m 和 σ 是 $\mathbf{R}\times(0,\infty)$上的实值函数且 $\sigma>0$，$\{W_t;\ t\geqslant 0\}$是独立于 X_0 的标准维纳过程。如前所述，式(7.4.1)中的记号是下述积分方程的简写形式

$$X_t = X_0 + \int_0^t m(X_s,s)\mathrm{d}s + \int_0^t \sigma(X_s,s)\mathrm{d}W_s \tag{7.4.2}$$

其中，第二个积分式可以解释为 Ito 随机积分。注意一个特例，即当 $m(X_t,\ t)=A_tX_t$ 和 $\sigma(X_t,\ t)=B_t$时，A_t 和 B_t 是某个确定性函数，式(7.4.1)是线性的。因此式(7.4.1)是 Kalman-Bucy 问题中提出的线性随机系统模型的推广形式。为了表述方便，本节所讨论的模型只局限于标量情况，向量的情况是类似的。

只有存在某个随机过程使得式(7.4.1)或式(7.4.2)成立，这两个表达式才有意义。幸运的是，只要函数 m 和 σ 满足简单的光滑性和递增性条件就能确保该式存在唯一的良性解。假设$E\{X_0^2\}<\infty$，那么存在上述随机过程的充分条件是该过程满足线性增长条件，即

$$|m(x,t)| + |\sigma(x,t)| \leqslant K[1+x^2]^{1/2} \tag{7.4.3}$$

其中 $K>0$。则该随机过程唯一的充分条件是满足一致 Lipschitz 条件，即

$$|m(x,t)-m(y,t)| + |\sigma(x,t)-\sigma(y,t)| \leqslant K|x-y| \tag{7.4.4}$$

其中 $K>0$ 且对 $\mathbf{R}^2$ 上的所有$(x,\ y)$式(7.1.4)均成立，因此，在下文中，完全可以用式(7.4.2)来定义一个连续采样的唯一随机过程。在这些过程中，$m(X_t,\ t)\mathrm{d}t$ 通常可以看成是偏移项，而 $\sigma(X_t,\ t)\mathrm{d}W_t$ 对应方差项。

形如式(7.4.1)的过程称之为扩散过程。该过程是马尔可夫过程，即对于每一个 $s>0$，给定 X_s 后，$\{X_t;\ t>s\}$和$\{X_t;\ 0\leqslant t<s\}$相互独立。可以直观地看出，这个性质对于任何的 $t\geqslant s\geqslant 0$ 都成立，并有如下表达式成立

$$X_t = X_s + \int_s^t m(X_u,u)\mathrm{d}u + \int_s^t \sigma(X_u,u)\mathrm{d}W_u \tag{7.4.5}$$

式(7.4.5)表明$\{X_t;\ t>s\}$可以完全由X_s和$\{W_t-W_s;\ t>s\}$构造。因此，只要给定X_s，由于$\{W_t-W_s;\ t>s\}$与过去的所有观测无关，则$\{X_t;\ t>s\}$与$\{X_t;\ t<s\}$无关。

满足马尔可夫性质的随机过程构成了一大类在实际中经常使用的随机模型，它比扩散过程的应用范围还要广。不过，在满足某些正则条件下，所有连续的马尔可夫过程都能表示成式(7.4.1)所描述的扩散过程的形式，其偏移项由下式给出

$$m(x,t) \triangleq \lim_{\Delta\to 0^+}\frac{E\{X_{t+\Delta}\,|\,X_t=x\}-x}{\Delta} \tag{7.4.6}$$

方差项则由下式的正平方根给出

$$\sigma^2(x,t) \triangleq \lim_{\Delta\to 0^+}\frac{\mathrm{var}\{X_{t+\Delta}\,|\,X_t=x\}}{\Delta} \tag{7.4.7}$$

因此，由式(7.4.1)所描述的扩散过程(以及与其相对的向量形式)能够包含大部分实际模型。[注意，在式(7.4.6)和式(7.4.7)中的$m(x,\ t)$和$\sigma^2(x,\ t)$分别表示对于给定$X_t=x$时，随机过程条件均值和方差对时间的导数。]

相比一般的非高斯随机过程而言，在马尔可夫性质的约束下，描述扩散随机过程的统计特性就更为容易。特别地，对于任意正整数n和n个时刻$0\leqslant t_1<t_2<\cdots<t_{n-1}<t_n<\infty$，可将$X_{t_1}$，…，$X_{t_n}$的联合概率密度分布表示为[参见式(3.2.99)]

$$p_{X_{t_1},\cdots,X_{t_n}}(x_1,\cdots,x_n)=p_{X_{t_1}}(x_1)\prod_{k=2}^{n}p_{X_{t_k}|X_{t_{k-1}}}(x_k\,|\,x_{k-1}) \tag{7.4.8}$$

因此，随机过程的任意有限维分布都可以根据初始条件的概率密度p_{X_0}和转移概率密度$p_{X_t|X_s}(x\,|\,y)$，$t>s\geqslant 0$计算。而且，当具有正则性时[⊖]，转移概率密度函数满足偏微分方程

$$\frac{\partial}{\partial t}p_{X_t|X_s}(x\,|\,y)=\mathcal{A}_t^*\,p_{X_t|X_s}(x\,|\,y),\quad t>s \tag{7.4.9}$$

其中，对于函数$f(x)$，$\mathcal{A}_t^*f(x)$表示算子

$$\mathcal{A}_t^*f(x)=\frac{1}{2}\frac{\partial^2}{\partial x^2}[\sigma^2(x,t)f(x)]-\frac{\partial}{\partial x}[m(x,t)f(x)] \tag{7.4.10}$$

式(7.4.9)就是著名的 Fokker-Planck 方程，也是扩散过程的 Kolmogorov 前向扩散方程。

Fokker-Planck 方程的一个重要推论是随机过程的一阶概率密度满足偏微分方程

$$\frac{\partial}{\partial t}p_{X_t}(x)=\mathcal{A}_t^*\,p_{X_t}(x),\quad t>0,\quad x\in\mathbf{R} \tag{7.4.11}$$

在$t=0$时的边界条件由初始概率密度$p_{X_0}(x)$给出。从式(7.4.11)可以得到形如$f(X_t)$的随机变量的期望的微分方程，其中f是$\mathbf{R}$上的实值函数。特别地，注意

$$E\{f(X_t)\}=\int_{-\infty}^{\infty}f(x)\,p_{X_t}(x)\,\mathrm{d}x \tag{7.4.12}$$

因此，假设积分和微分的顺序可以交换，那么有

$$\frac{\mathrm{d}}{\mathrm{d}t}E\{f(X_t)\}=\int_{-\infty}^{\infty}f(x)\,\frac{\partial}{\partial t}p_{X_t}(x)\,\mathrm{d}x=\int_{-\infty}^{\infty}f(x)\,\mathcal{A}_t^*\,p_{X_t}(x)\,\mathrm{d}x \tag{7.4.13}$$

⊖ 一组充分条件是$\sigma(x,\ t)$永不为0且$\sigma(x,\ t)$、$m(x,\ t)$、$\partial\sigma(x,\ t)/\partial x$、$\partial m(x,\ t)/\partial x$及$\partial^2\sigma(x,\ t)/\partial x^2$都满足线性递增和一致 Lipschitz 条件。

进一步假设 f 是充分正则的[⊖]，通过分部积分后，式(7.4.13)简化为

$$\frac{\mathrm{d}}{\mathrm{d}t}E\{f(X_t)\} = E\{\mathcal{A}_t f(X_t)\}, \quad t \geqslant 0 \tag{7.4.14}$$

初始条件为 $E\{f(X_0)\}$，运算$\mathcal{A}_t$ 定义为

$$\mathcal{A}_t f(X_t) \triangleq m(x,t)\frac{\mathrm{d}}{\mathrm{d}x}f(x) + \frac{1}{2}\sigma^2(x,t)\frac{\mathrm{d}^2}{\mathrm{d}x^2}f(x)$$

假设 X_t 的均值 $u_t = E\{X_t\}$，$f(x)=x$，根据式(7.4.14)，有

$$\dot{\mu}_t = E\{m(X_t,t)\}, \quad t \geqslant 0, \mu_0 = E\{X_0\} \tag{7.4.15}$$

其中，变量上方的圆点表示时间微分。类似地，X_t 的方差 $v_t = E\{X_t^2\} - \mu_t^2$ 满足微分方程

$$\dot{v}_t = \frac{\mathrm{d}}{\mathrm{d}t}E\{X_t^2\} - 2\mu_t\dot{\mu}_t = 2\mathrm{cov}(X_t, m(X_t,t)) + E\{\sigma^2(X_t,t)\}, \quad t \geqslant 0 \tag{7.4.16}$$

且 $v_0 = \mathrm{var}(X_0)$。要求解上述两个等式通常较为困难，这是因为在所求的期望中包含 $m(X_t, t)$和 $\sigma^2(X_t, t)$。虽然这两个量满足形如式(7.4.14)的微分方程，但是这些方程涉及求关于 X_t 的非线性函数的期望。可以看出，在非线性滤波问题中也存在类似的困难。

线性情形则是此类问题的一个简化求解，此时，$m(X_t, t) = A_t X_t$，$\sigma(X_t, t) = B_t$，其中 A_t 和 B_t 都是确知函数。进一步可将式(7.4.15)和式(7.4.16)简化为

$$\dot{\mu}_t = A_t\mu_t, \quad t \geqslant 0, \quad \mu_0 = E\{X_0\} \tag{7.4.17}$$

和

$$\dot{v}_t = 2A_t + B_t^2, \quad t \geqslant 0, \quad v_0 = \mathrm{var}(X_0) \tag{7.4.18}$$

式(7.4.17)和式(7.4.18)是不带其他未知量的一阶线性微分方程。事实上，式(7.4.17)和式(7.4.18)分别对应不含观测量($C_t = 0$，$t \geqslant 0$)时卡尔曼滤波方程的均值和方差。

注意，不需要推导 Fokker-Planck 方程的中间步骤，描述 $E\{f(X_t)\}$演化的微分方程(7.4.14)也可以根据扩散方程(7.4.1)直接得到。特别地，假设 f 有连续的一阶导数 f' 和二阶导数 f''，那么可以利用泰勒展开理论，在 $\Delta > 0$ 时，将式(7.4.1)改写成

$$\begin{aligned}
\frac{f(X_{t+\Delta}) - f(X_t)}{\Delta} &\sim \frac{X_{t+\Delta} - X_t}{\Delta}f'(X_t) + \frac{1}{2}\frac{(X_{t+\Delta} - X_t)^2}{\Delta}f''(X_t) \\
&\sim \left[m(X_t,t) + \sigma(X_t,t)\frac{W_{t+\Delta} - W_t}{\Delta}\right]f'(X_t) \\
&\quad + \frac{1}{2}\frac{[m(X_t,t)\Delta + \sigma(X_t,t)(W_{t+\Delta} - W_t)]^2}{\Delta}f''(X_t) \\
&= m(X_t,t)f'(X_t) + \sigma(X_t,t)f'(X_t)\frac{W_{t+\Delta} - W_t}{\Delta} \\
&\quad + \frac{\Delta}{2}m^2(X_t,t)f''(X_t) + m(X_t,t)\sigma(X_t,t)f''(X_t)(W_{t+\Delta} - W_t)
\end{aligned}$$

⊖ 充要条件是 $\partial^2 f(x)/\partial x^2$ 存在且连续，并满足

$$\begin{aligned}
\lim_{|x|\to\infty} m(x, t)f(x)p_{X_t}(x) &= \lim_{|x|\to\infty}\left[\frac{\partial}{\partial x}f(x)\right]\left[\frac{\partial}{\partial x}\sigma(x, t)p_{X_t}(x)\right] \\
&= \lim_{|x|\to\infty}\left[\frac{\partial^2}{\partial x^2}f(x)\right][\sigma(x, t)p_{X_t}(x)] = 0
\end{aligned}$$

$$+\frac{1}{2}\sigma^2(X_t,t)f''(X_t)\frac{(W_{t+\Delta}-W_t)^2}{\Delta} \tag{7.4.19}$$

注意，$(W_{t+\Delta}-W_t)$服从$\mathcal{N}(0,\ \Delta)$分布且与 X_t 无关，则对式(7.4.19)两边同取期望可以得到

$$\frac{E\{f(X_{t+\Delta})\}-E\{f(X_t)\}}{\Delta}\sim E\{m(X_t,t)f'(X_t)\}+\frac{\Delta}{2}E\{m^2(X_t,t)f''(X_t)\}$$
$$+\frac{1}{2}E\{\sigma^2(X_t,t)f''(X_t)\}$$

假设 $E\{m^2(X_t,\ t)f''(X_t)\}<\infty$，且令 $\Delta\to 0$ 后，有下面的表达式成立

$$\frac{\mathrm{d}}{\mathrm{d}t}E\{f(X_t)\}=E\left\{m(X_t,t)f'(X_t)+\frac{1}{2}\sigma^2(X_t,t)f''(X_t)\right\}=E\{\mathcal{A}_tf(X_t)\}$$

即式(7.4.14)。

注意，$E\{f(X_t)\}=\int_{-\infty}^{\infty}f(x)p_{X_t}(x)\mathrm{d}x$，那么对每个给定的 x 而言，$p_{X_t}(x)$可视为 $E\{\delta(x-X_t)\}$，其中 δ 是狄拉克冲激函数。将期望的计算形式代入式(7.1.13)后，有

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{-\infty}^{\infty}f(x)p_{X_t}(x)\mathrm{d}x=\int_{-\infty}^{\infty}f(x)\ \mathcal{A}_t^*\,p_{X_t}(x)\mathrm{d}x$$

其中，令 f 为满足 $\partial p_{X_t}(x)/\partial x=\mathcal{A}_t^*\,p_{X_t}(x)$的脉冲响应。(事实上，Fokker-Planck 方程本身也是以 $X_s=y$ 时的期望为条件推导出来的。)虽然最后一步推导并不是那么严格，但是由于上面已经假设 f 有连续的二阶导数并可将狄拉克冲激函数看成是光滑函数的极限，所以对于那些足够正则的 m 和 σ 而言是严格的。因此，期望的微分方程和概率密度的微分方程包含了相同的扩散过程信息，这是推导非线性滤波的基本方程的关键。

7.4.1 非线性滤波的基本方程

现在考虑由式(7.4.1)产生的扩散过程$\{X_t;\ t\geqslant 0\}$通过含噪的非线性观测得到，即

$$Y_t=h(X_t,t)+N_t,\quad t\geqslant 0 \tag{7.4.20}$$

其中，h 是在 $\mathbf{R}\times(0,\ \infty)$上的实值函数，$\{N_t;\ t\geqslant 0\}$是独立于$\{W_t;\ t\geqslant 0\}$和 X_0 的零均值高斯过程且满足 $E\{N_tN_s\}=R_t\delta(t-s)$(对所有 $t\geqslant 0$ 都有 $R_t>0$)。为了得到更加严格的模型，对式(7.4.20)积分得到的观测量$\{Z_t;\ t\geqslant 0\}$由下式给出

$$\mathrm{d}Z_t=h(X_t,t)\mathrm{d}t+R_t^{1/2}\mathrm{d}V_t,\quad t\geqslant 0 \tag{7.4.21}$$

其中，$\{V_t;\ t\geqslant 0\}$是与 X_0 和$\{W_t;\ t\geqslant 0\}$无关的标准维纳过程。

目的是对于每一个 $t\geqslant 0$，利用从观察起始时刻直至时间 t 的观测量来估计 X_t 的函数 $f(X_t)$。由第 4 章的讨论可知，条件均值是最小均方误差估计量，即

$$\widehat{f(X_t)}=E\{f(X_t)\,|\,Z_0^t\} \tag{7.4.22}$$

注意，式(7.4.22)可以通过下式计算得到

$$\widehat{f(X_t)}=\int_{-\infty}^{\infty}f(x)q_{X_t}(x)\mathrm{d}x \tag{7.4.23}$$

其中，$q_{X_t}(x)$表示给定 Z_0^t 后 X_t 的条件概率密度函数。[为了简单起见，没有将其记为条件概率函数的表达式，但是要注意 $q_{X_t}(x)$是依赖于 Z_0^t 的函数。]注意，当缺乏观测量时[即 $h(X_t,\ t)\equiv 0$]，$q_{X_t}(x)$就是 $p_{X_t}(x)$，且$\widehat{f(X_t)}$满足微分方程(7.4.14)。因此，如果可以找

到某个方程使得条件概率密度函数 $q_{X_t}(x)$ 与 Fokker-Planck 方程中出现的概率密度函数 $p_{X_t}(x)$ 类似，那么就可以给出条件均值为式(7.4.22)的类似的微分方程。不过，对每个 $x\in\mathbf{R}$ 而言，$\{q_{X_t}(x);\ t\geqslant 0\}$ 是随机过程，因此这个方程只有在统计意义上才成立。

在正则条件下，可以证明这样的方程确实存在。特别地，$q_{X_t}(x)$ 满足如下的随机偏微分方程

$$\mathrm{d}q_{X_t}(x)=\mathcal{A}_t^{*}q_{X_t}(x)\mathrm{d}t+q_{X_t}(x)[h(x,t)-\widehat{h(X_t,t)}]R_t^{-1}\mathrm{d}I_t,\quad t\geqslant 0 \tag{7.4.24}$$

其中，此处及后文出现的带“三角尖”的 X_t 的函数表示给定 Z_0^t 的条件均值。式(7.4.24)中的过程 $\{I_t;\ t\geqslant 0\}$ 是(非线性)新息过程，即

$$I_t\triangleq Z_t-\int_0^t\widehat{h(X_s,s)}\mathrm{d}s,\quad t\geqslant 0 \tag{7.4.25}$$

其初始条件是 $q_{X_0}(x)=p_{X_0}(x)$。式(7.4.24)也称为 Kushner 方程。注意，对于 p_{X_t} 而言，当没有观测量(即 $h\equiv 0$)时，该方程将退化为 FokkerPlanck 方程。否则，条件概率密度的演化将由系统的动力学特性(通过运算符 $\mathcal{A}_t^{*}$)和观测量共同描述。式(7.4.24)中的这组方程给出了 $x\in\mathbf{R}$ 的概率密度 $q_{X_t}(x)$ 的递归形式，因为方程的右边只包含了 $q_{X_t}(x)$、$q_{X_t}(x)$ 的导数和可以根据 $q_{X_t}(x)$ 计算的 $\widehat{h(X_t,\ t)}$。式(7.4.24)的具体推导过程将会在后面给出。

结合式(7.4.23)和式(7.4.24)，可以得到由 $\widehat{f(X_t)}$ 演化的随机微分方程。特别地，首先给出式(7.4.24)的积分形式，即

$$q_{X_t}(x)=p_{X_0}(x)+\int_0^t q_{X_s}(x)\mathrm{d}s+\int_0^t\mathcal{A}_s^{*}q_{X_s}(x)[h(x,s)-\widehat{h(X_s,s)}]R_s^{-1}\mathrm{d}I_s \tag{7.4.26}$$

假设积分的顺序是可以交换的，则有

$$\begin{aligned}\widehat{f(X_t)}&=\int_{-\infty}^{\infty}f(x)q_{X_t}(x)\mathrm{d}x=\int_{-\infty}^{\infty}f(x)p_{X_0}(x)\mathrm{d}x+\int_0^t\int_{-\infty}^{\infty}f(x)\,\mathcal{A}_s^{*}q_{X_s}(x)\mathrm{d}x\\&\quad+\int_0^t\int_{-\infty}^{\infty}f(x)q_{X_s}(x)[h(x,s)-\widehat{h(X_s,s)}]\mathrm{d}x\mathrm{d}I_s\\&=E\{f(X_0)\}+\int_0^t\int_{-\infty}^{\infty}q_{X_s}(x)\,\mathcal{A}_sf(x)\mathrm{d}x\mathrm{d}s\\&\quad+\int_0^t\left[\int_{-\infty}^{\infty}f(x)h(x,s)q_{X_s}(x)\mathrm{d}x-\int_{-\infty}^{\infty}f(x)q_{X_s}(x)\mathrm{d}x\,\widehat{h(X_s,s)}\right]R_s^{-1}\mathrm{d}I_s\\&=E\{f(X_0)\}+\int_0^t\widehat{\mathcal{A}_sf(X_s)}\mathrm{d}s+\int_0^t[\widehat{f(X_s)h(X_s,s)}-\widehat{f(X_s)}\,\widehat{h(X_s,s)}]R_s^{-1}\mathrm{d}I_s\end{aligned} \tag{7.4.27}$$

其中，在第二个等式中，利用了式(7.4.15)用过的分部积分法。式(7.4.27)可以重写为

$$\begin{aligned}&\mathrm{d}\,\widehat{f(X_t)}=\widehat{\mathcal{A}_tf(X_t)}\mathrm{d}t+[\widehat{f(X_t)h(X_t,t)}-\widehat{f(X_t)}\,\widehat{h(X_t,t)}]R_t^{-1}\mathrm{d}I_t,\quad t>0\\&\widehat{f(X_0)}=E\{f(X_0)\}\end{aligned} \tag{7.4.28}$$

上述方程中的 $\widehat{f(X_t)h(X_t,\ t)}-\widehat{f(X_t)}\widehat{h(X_t,\ t)}$ 项是给定 Z_0^t 时，函数 $f(X_t)$ 和 $h(X_t,\ t)$ 的条件方差。

在 $f(X_t)$ 依赖于时间 t 时，可以得出和式(7.4.28)类似的方程。在考虑 $f(X_t, t)$ 后，方程(7.4.28)唯一的改变是在右边需要增加 $\widehat{\dot{f}(X_t, t)}\mathrm{d}t$ 项，其中 $\dot{f}=\partial f/\partial t$。注意，当没有观测量($h\equiv 0$)时，在式(7.4.28)中，令 $f(X_t)=\delta(x-X_t)$ 就能得到计算条件概率密度 $q_{X_t}(x)$ 所需的式(7.4.24)。因此，式(7.4.28)和式(7.4.24)在本质上等价。

虽然式(7.4.28)描述的是从已知条件 Z_0^t 中估计 $f(X_t)$ 的滤波器的演化过程，但是该方程并没有指出如何应用该滤波器。这是因为式(7.4.28)的实现同样也需要求解 $\widehat{\mathcal{A}_t^* f}$、$\widehat{fh}$ 和 $\hat{h}$ 的滤波器，而这又会涉及求解更深层次的非线性滤波器。除非在有限次的应用式(7.4.28)之后，包含在这些方程中的变量能够得到闭式解，否则通过这种方法推导出的关于 $\widehat{f(X_t)}$ 的滤波器将会是无限维的。相反，如上所述，虽然条件概率密度的方程严格递归，但是因为条件概率密度函数是值函数，只有在无限维参数情况下才是递归的。换言之，只有在一些特殊的情形下才能有效求解。

在下面的例子中将讨论这些情形和滤波方程(7.4.28)的其他方面。尽管从滤波器实现的角度看，最优滤波方程一般并不合适，但是在对最优滤波器进行近似时却非常有用，并且能够得到在实际中可行的滤波器。下面，通过例子来对式(7.4.28)近似后的一些实用滤波器进行简要讨论。

例 7.4.1　直接状态估计之条件均值演化过程

所感兴趣的通常是直接估计扩散项 X_t 本身。与之相对应的情形是 $f(x)=x$，滤波方程(7.4.28)变成

$$\mathrm{d}\hat{X}_t=\widehat{m(X_t,t)}\mathrm{d}t+[\widehat{X_t h(X_t,t)}-\hat{X}_t\widehat{h(X_t,t)}]R_t^{-1}\mathrm{d}I_t \tag{7.4.29}$$

注意，没有观测量($h\equiv 0$)时式(7.4.29)将退化成式(7.4.15)。

对于线性观测量[$h(X_t, t)=C_t X_t$]，式(7.4.29)可以写为

$$\mathrm{d}\hat{X}_t=\widehat{m(X_t,t)}\mathrm{d}t+C_t[\widehat{X_t^2}-\hat{X}_t^2]R_t^{-1}\mathrm{d}I_t=\widehat{m(X_t,t)}\mathrm{d}t+C_tP_tR_t^{-1}\mathrm{d}I_t \tag{7.4.30}$$

其中，$P_t\triangleq\widehat{X_t^2}-\hat{X}_t$ 是给定 Z_0^t 时，X_t 的条件方差，即条件均方误差(MSE)。可以发现，更新项 $C_tP_tR_t^{-1}\mathrm{d}I_t$ 与卡尔曼滤波器中相对应项的形式一样[参见式(7.3.61)]。进而，如果假设偏移项也具有线性性质，即 $m(X_t, t)=A_tX_t$，则滤波方程(7.4.30)退化为

$$\mathrm{d}\hat{X}_t=A_t\hat{X}_t\mathrm{d}t+C_tP_tR_t^{-1}\mathrm{d}I_t \tag{7.4.31}$$

式(7.4.31)与 Kalman-Bucy 模型中出现的滤波方程形式一致。然而，这两者之间存在着根本差异，即式(7.4.31)中的 P_t 可能依赖于 Z_0^t。下面的例子将考虑这一问题。◀

例 7.4.2　直接状态估计之条件方差演化过程

例 7.4.1 中对 $\hat{X}_t$ 的估计性能可以用上面介绍的条件均方误差(MSE) $P_t=\widehat{X_t^2}-\hat{X}_t^2$ 来衡量。可以通过 $\mathrm{d}P_t=\mathrm{d}\widehat{X_t^2}-\mathrm{d}(\hat{X}_t^2)$ 来研究估计器的性能，其中 $\mathrm{d}\widehat{X_t^2}$ 对应 $f(x)=x^2$ 的情况。显然，直接应用式(7.4.28)后，有

$$\mathrm{d}\hat{X}_t^2=2\widehat{X_t m(X_t,t)}\mathrm{d}t+\widehat{\sigma^2(X_t,t)}\mathrm{d}t+[\widehat{X_t^2 h(X_t,t)}-\widehat{X_t^2}\widehat{h(X_t,t)}]R_t^{-1}\mathrm{d}I_t \tag{7.4.32}$$

对于第二项 $d(\hat{X}_t^2)$，原则上可以将其写为 $d(\hat{X}_t^2)=2\hat{X}_t d\hat{X}_t$，并利用 $\hat{X}_t$ 的演化方程来消去 $d\hat{X}_t$。不过，式(7.4.30)中的 $\hat{X}_t$ 包含关于更新过程 $\{I_s;\ 0\leqslant s\leqslant t\}$ 的 Ito 随机积分，而根据 6.4 节(参见 6.4.5 节)已经知道这类过程一般不满足 $\int_0^t X_s dX_s=\frac{1}{2}(X_t^2-X_0^2)$，所以需要额外考虑。

为了推导出 $d(\hat{X}_t^2)$ 的正确表达式，首先给出 Ito 微分规则或 Ito 公式。 ◀

命题 7.4.1　Ito 微分规则

假设 F_t 是下列形式的随机过程

$$dF_t=G_t dt+H_t dW_t,\quad 0\leqslant t\leqslant T \tag{7.4.33}$$

其中 G_0^T 和 H_0^T 是满足 $P\left(\int_0^T|G_t|dt<\infty\right)=P\left(\int_0^T|H_t^2|dt<\infty\right)=1$ 的随机过程，W_0^T 是维纳过程，且对于每一个 $s\in(0,\ T)$，都有 $\{W_t-W_s;\ s\leqslant t\leqslant T\}$ 和 $(\mathcal{G}_0^s,\ H_0^s,\ W_0^s)$ 独立。进一步假设 g 是在 $\mathbf{R}$ 上的实值函数且有连续的一阶和二阶导数 g' 和 g''。定义 $J_t=g(F)$，有

$$dJ_t=g'(F_t)dF_t+\frac{1}{2}g''(F_t)H_t^2,\quad 0\leqslant t\leqslant T \tag{7.4.34}$$

备注

1) 在普通微积分中与式(7.4.33)相对应的公式是 $dJ_t=g'(F_t)dF_t$。出现修正项 $(1/2)g''(F_t)H_t^2$ 是因为：对于维纳过程而言，从根本上说，$(dW_t)^2$ 的作用更像 dt 而不是 $(dt)^2$。对于形如式(7.4.33)的这些过程来说，将其理解成 $(dF_t)^2\sim H_t^2 dt$。特别地，作为对式(7.4.34)的一个合理的论证，对 $dg(F_t)$ 可以利用泰勒展开式得到

$$\begin{aligned}dg(F_t)&\sim g(F_{t+dt})-g(F_t)\sim g'(F_t)(F_{t+dt}-F_t)+\frac{1}{2}g''(F_t)(F_{t+dt}-F_t)^2\\&\sim g'(F_t)dF_t+\frac{1}{2}g''(F_t)(dF_t)^2\end{aligned} \tag{7.4.35}$$

利用式(7.4.33)，有

$$(dF_t)^2=G_t^2(dt)^2+2G_tH_t(dW_t)dt+H_t^2(dW_t)^2$$

该式中的第一项是 $O((dt)^2)$，第二项是 $O((dt)^{3/2})$，第三项是 $\sim H_t^2 dt$。因此，如上所述，有 $(dF_t)^2\sim H_t^2 dt$ 成立，进而得到式(7.4.34)。Lipster 和 Shiryayev(1977)对该结果给出了严格证明。

2) 与式(6.4.4)一样，式(7.4.35)的特例是 $g(x)=x^2$，$G_t=0$ 而且 H_t 是一个确知函数(称为 ϕ_t)。鉴于满足命题假设的所有 g 都是局部二次的，因而 $g(x)=x^2$ 是典型例子。

3) 结合式(7.4.33)和式(7.4.34)，可以得到

$$dJ_t=[g'(F_t)G_t+\frac{1}{2}g''(F_t)H_t^2]dt+g'(F_t)H_t dW_t \tag{7.4.36}$$

因此 J_t 也是形如式(7.4.33)的过程，其中需要将 G_t 替换成 $[g'(F_t)G_t+(1/2)g''(F_t)H_t^2]$，$H_t$ 替换成 $g'(F_t)H_t$。可以看出，形如式(7.4.34)所描述的一类过程在光滑的(无记忆的)非线性变换下是封闭的。此类过程(包括扩散过程)有时也称为 Ito 过程。

现在让回到之前讨论的条件方差 P_t 的演化问题。在对 $(\hat{X}_t)^2$ 应用 Ito 公式之前，必须

先将 $\mathrm{d}\hat{X}_t$ 写成式(7.4.33)的形式。注意

$$\mathrm{d}I_t = \mathrm{d}Z_t - \widehat{h(X_t,t)}\mathrm{d}t = [h(X_t,t) - \widehat{h(X_t,t)}]\mathrm{d}t + R_t^{1/2}\mathrm{d}V_t$$

$\hat{X}_t$ 的滤波方程变为

$$\mathrm{d}\hat{X}_t = \{\widehat{m(X_t,t)} + [\widehat{X_t h(X_t,t)} - \hat{X}_t\widehat{h(X_t,t)}] \times R_t^{-1}[h(X_t,t) - \widehat{h(X_t,t)}]\}\mathrm{d}t$$
$$+ [\widehat{X_t h(X_t,t)} - \hat{X}_t\widehat{h(X_t,t)}]R_t^{-1/2}\mathrm{d}V_t$$

由于 V_t 是维纳过程且$\{V_t - V_s;\ t \geqslant s\}$独立于过去的时刻，应用 Ito 公式后可得到

$$\mathrm{d}\hat{X}_t^2 = 2\hat{X}_t\mathrm{d}\hat{X}_t + [\widehat{X_t h(X_t,t)} - \hat{X}_t\widehat{h(X_t,t)}]^2 R_t^{-1}\mathrm{d}t \tag{7.4.37}$$

联立式(7.4.30)、式(7.4.32)和式(7.4.37)并且重排各项，得到 X_t 的条件方差的随机微分方程如下

$$\mathrm{d}P_t = \{2\widehat{(X_t - \hat{X}_t)m(X_t,t)} + \widehat{\sigma^2(X_t,t)} - [\widehat{X_t h(X_t,t)} - \hat{X}_t\widehat{h(X_t,t)}]^2 R_t^{-1}\}\mathrm{d}t$$
$$+ [\widehat{(X_t - \hat{X}_t)^2 h(X_t,t)} - P_t\widehat{h(X_t,t)}]R_t^{-1}\mathrm{d}I_t,\quad t > 0 \tag{7.4.38}$$

其中 $P_0 = \mathrm{var}(X_0)$。

一般来说，该方程非常复杂。对于线性观测$[h(X_t,\ t) = C_t X_t]$而言，则可以简化为

$$\mathrm{d}P_t = [2\widehat{(X_t - \hat{X}_t)m(X_t,t)} + \widehat{\sigma^2(X_t,t)} + C_t^2 P_t^2 R_t^{-1}]\mathrm{d}t + \widehat{(X_t - \hat{X}_t)^3}C_t R_t^{-1}\mathrm{d}I_t \tag{7.4.39}$$

注意，式(7.4.39)右边的二次项 $C_t^2 P_t^2 R_t^{-1}$，在 Kalman-Bucy 模型中，该项在由式(7.3.62)所描述的 Riccati 方程中也出现过，在那里它表示误差方差。如果假设偏移项也是线性的$[m(X_t,\ t) = A_t X_t]$，则式(7.4.39)可以进一步简化为

$$\mathrm{d}P_t = [2A_t P_t + \widehat{\sigma^2(X_t,t)} + C_t^2 P_t^2 R_t^{-1}]\mathrm{d}t + \widehat{(X_t - \hat{X}_t)^3}C_t R_t^{-1}\mathrm{d}I_t \tag{7.4.40}$$

(一次项 $2A_t P_t$ 同样也出现在 Kalman-Bucy 的方差方程中)注意，尽管在线性观测和偏移假设下的滤波方程(7.4.31)与 Kalman-Bucy 滤波方程具有相同的形式，但根据式(7.4.40)，它们存在重要区别，即式(7.4.31)中的方差 P_t 需要满足式(7.4.40)。换言之，它依赖于观测量。

假设 $\sigma(x,\ t)$不依赖于 x，例如 $\sigma(x,\ t) = B_t$，式(7.4.40)可以进一步简化为

$$\mathrm{d}P_t = (2A_t P_t + B_t^2 + C_t^2 P_t^2 R_t^{-1})\mathrm{d}t + \widehat{(X_t - \hat{X}_t)^3}R_t^{-1}\mathrm{d}I_t \tag{7.4.41}$$

因此，即便对于线性模型，条件方差 P_t 也仍然依赖于观测，这是由于更新项中包含了条件三阶中心矩$\widehat{(X_t - \hat{X}_t)^3}$。然而，如果假设初始条件 X_0 是高斯过程，那么该模型中的所有过程都是联合高斯过程，即给定 Z_0^t 后，X_t 也满足条件高斯分布。由于高斯分布关于其均值对称，奇数阶中心距均为零，所以有$\widehat{(X_t - \hat{X}_t)^3} = 0$ 和

$$\mathrm{d}P_t = (2A_t P_t + B_t^2 + C_t^2 P_t^2 R_t^{-1})\mathrm{d}t \tag{7.4.42}$$

式(7.4.42)为 Kalman-Bucy 滤波中的方差方程。

7.4.2 非线性滤波方程的推导

现在给出式(7.4.28)的简单推导过程。表达式为

$$\mathrm{d}\widehat{f(X_t)} = \widehat{\mathcal{A}_t f(X_t)}\mathrm{d}t + [\widehat{f(X_t)h(X_t,t)} - \widehat{f(X_t)}\,\widehat{h(X_t,t)}]R_t^{-1}\mathrm{d}I_t$$

上式描述了在给定 Z_0^t 的条件下 $f(X_t)$ 最小均方误差估计器的演化过程。正如上面所提到的，令 $f(X_t)=\sigma(x-X_t)$ 将得到描述条件密度演化过程的式(7.4.24)。在推导式(7.4.28)的过程中，将得到与概率密度式(7.4.24)具有相同信息的方程，但是形式却更加简单。为了简单起见，假设对所有的 $t\geqslant 0$ 都有 $R_t=1$。对于一般的 R_t 来说可以进行直接修正。

需要找到条件期望 $E\{f(X_t)\mid Z_0^t\}$ 的某种表达式。为此先考虑给定 X_0^t 下 Z_0^t 的条件统计特性，然后利用贝叶斯公式求解条件期望。如果知道扩散过程在 t 时刻及之前所有时刻的样本，假定 $X_0^t=x_0^t$ 后可将 Z_0^t 视为加性维纳过程中的已知波形，即

$$\mathrm{d}Z_s = h(x_s,s)\mathrm{d}s + \mathrm{d}V_s,\quad 0\leqslant s\leqslant t \tag{7.4.43}$$

在假设 $\int_0^t h^2(x_s,s)\mathrm{d}s<\infty$ 后(如果 h 是有界的，那么该条件总是成立的，且如果 h 对于这两个变量来说是连续的，那么该条件依概率 1 成立)，就可以在维纳意义上给出 Z_0^t 的概率密度函数。特别地，该概率密度函数将会是假设 Z_0^t 由式(7.4.43)给出和假设 Z_0^t 是标准维纳过程的似然比。将该似然比记为 Λ_t，其表达式由 Cameron-fMartin 公式给出，即

$$\Lambda_t(x_0^t;Z_0^t) = \exp\left\{\int_0^t h(x_s,s)\mathrm{d}Z_s - \frac{1}{2}\int_0^t h^2(x_s,s)\mathrm{d}s\right\}$$

令 $P_{X_0^t}$ 表示与 X_0^t 有关的概率测度[⊖]。再根据贝叶斯公式，有

$$\widehat{f(X_t)} = \frac{\int f(X_t)\Lambda_t(x_0^t;Z_0^t)P_{X_0^t}(\mathrm{d}x_0^t)}{\int \Lambda_t(x_0^t;Z_0^t)P_{X_0^t}(\mathrm{d}x_0^t)} \triangleq \frac{N_t(Z_0^t)}{D_t(Z_0^t)} \tag{7.4.44}$$

其中 $N_t(Z_0^t)$ 和 $D_t(Z_0^t)$ 分别记为该等式中间项的分子和分母。

推导关于 $f(X_t)$ 演化的方程，首先要找到关于 $N_t(Z_0^t)$ 和 $D_t(Z_0^t)$ 的方程，然后将其结合。为此，下面将给出向量过程的 Ito 微分公式。

命题 7.4.2　向量 Ito 公式

假设 $\{W_t;\ t\geqslant 0\}$ 是 k 维独立的标准向量维纳过程，且 $\{F_t;\ t\geqslant 0\}$ 是由下列方程产生的 m 维随机向量

$$\mathrm{d}\boldsymbol{F}_t = \boldsymbol{G}_t\mathrm{d}t + \boldsymbol{H}_t\mathrm{d}\boldsymbol{W}_t,\quad 0\leqslant t\leqslant T \tag{7.4.45}$$

其中 $\boldsymbol{G}_t$ 是 m 维随机向量，$\boldsymbol{H}_t$ 是 $m\times n$ 的随机过程矩阵。对于所有的 $s\geqslant 0$ 来说，$\{\boldsymbol{W}_t-\boldsymbol{W}_s;\ t\geqslant s\}$ 独立于($\boldsymbol{G}_0^s$，$\boldsymbol{H}_0^s$，$\boldsymbol{W}_0^s$)。$\boldsymbol{G}_t$ 的每个分量和 $\boldsymbol{H}_t$ 的每个元素在[0，T]上依概率 1 绝对可积。[式(7.4.45)表明 $\boldsymbol{F}_t$ 的第 j 个分量可以写为 $F_t^{(j)} = F_0^{(j)} + \int_0^t G_s^{(j)}\mathrm{d}s + \sum_{k=1}^m\int_0^t H_s^{(j,k)}\mathrm{d}W_s^{(k)}$，$j=1,2,\cdots,m$]。进一步假定 g 是 $\mathbf{R}^m$ 上的实值函数且存在连续的二阶偏导数，则

$$\mathrm{d}g(\boldsymbol{F}_t) = \sum_{j=1}^m g_{x_j}(\boldsymbol{F}_t)\mathrm{d}F_t^{(j)} + \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m g_{x_i,x_j}(\boldsymbol{F}_t)\mathrm{d}\langle F^{(i)},F^{(j)}\rangle_t \tag{7.4.46}$$

其中，$g_{x_i}(\boldsymbol{x})=\partial g(\boldsymbol{x})/\partial x_i$，$g_{x_i,x_j}(\boldsymbol{x})=\partial^2 g(\boldsymbol{x})/\partial x_i\partial x_j$，且

$$\mathrm{d}\langle F^{(i)},\ F^{(j)}\rangle_t = (\boldsymbol{H}_t\boldsymbol{H}_t^T)_{i,j}\mathrm{d}t \tag{7.4.47}$$

其中 $\langle F^{(i)},\ F^{(j)}\rangle_0=0$。

⊖ 注意，X_0^t 和 Z_0^t 是[0，t]上的连续函数空间中的随机元素，满足 σ 代数性质[参见 Dunford 和 Schwartz (1958)]。测度 $P_{X_0^t}$ 和那些涉及定义 Λ_t 的测度都在该空间中。

备注

1）以向量泰勒级数的形式展开 $\mathrm{d}g(\boldsymbol{F}_t)\sim g(\boldsymbol{F}_{t+\mathrm{d}t})-g(\boldsymbol{F}_t)$后，证明此结论的过程与证明命题 7.4.1 类似。如果将式(7.4.46)重写为如下形式则会更加明显

$$\begin{aligned}\mathrm{d}g(\boldsymbol{F}_t)&=[\nabla g(\boldsymbol{F}_t)]^{\mathrm{T}}\mathrm{d}\boldsymbol{F}_t+\frac{1}{2}\mathrm{tr}\{\boldsymbol{Dg}(\boldsymbol{F}_t)\boldsymbol{H}_t\boldsymbol{H}_t^{\mathrm{T}}\}\mathrm{d}t\\&=[\nabla g(\boldsymbol{F}_t)]^{\mathrm{T}}\mathrm{d}\boldsymbol{F}_t+\frac{1}{2}\mathrm{tr}\{\boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{Dg}(\boldsymbol{F}_t)\boldsymbol{H}_t^{\mathrm{T}}\}\mathrm{d}t\end{aligned}\tag{7.4.48}$$

其中，∇g 表示 g 的梯度，$\boldsymbol{Dg}$ 是 g 的二阶偏导矩阵，$\mathrm{tr}\{\cdot\}$表示取迹操作（即 $\mathrm{tr}\boldsymbol{A}=\sum_{i=1}^{n}A_{i,i}$）。

展开后可得

$$\begin{aligned}\mathrm{d}g(\boldsymbol{F}_t)\sim g(\boldsymbol{F}_{t+\mathrm{d}t})-g(\boldsymbol{F}_t)&\sim[\nabla g(\boldsymbol{F}_t)]^{\mathrm{T}}(\boldsymbol{F}_{t+\mathrm{d}t}-\boldsymbol{F}_t)+\frac{1}{2}(\boldsymbol{F}_{t+\mathrm{d}t}-\boldsymbol{F}_t)^{\mathrm{T}}\boldsymbol{Dg}(\boldsymbol{F}_t)(\boldsymbol{F}_{t+\mathrm{d}t}-\boldsymbol{F}_t)\\&\sim[\nabla g(\boldsymbol{F}_t)]^{\mathrm{T}}\mathrm{d}\boldsymbol{F}_t+\frac{1}{2}\mathrm{d}\boldsymbol{F}_t^{\mathrm{T}}\boldsymbol{Dg}(\boldsymbol{F}_t)\mathrm{d}\boldsymbol{F}_t\end{aligned}$$

利用式(7.4.45)，可以得到

$$\begin{aligned}\mathrm{d}g(\boldsymbol{F}_t)\sim{}&[\nabla g(\boldsymbol{F}_t)]^{\mathrm{T}}\mathrm{d}\boldsymbol{F}_t+\frac{1}{2}\boldsymbol{G}_t^{\mathrm{T}}\boldsymbol{Dg}(\boldsymbol{F}_t)\boldsymbol{G}_t(\mathrm{d}t)^2+\boldsymbol{G}_t^{\mathrm{T}}\boldsymbol{Dg}(\boldsymbol{F}_t)\boldsymbol{H}_t(\mathrm{d}\boldsymbol{W}_t)\mathrm{d}t\\&+\frac{1}{2}(\mathrm{d}\boldsymbol{W}_t)^{\mathrm{T}}\boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{Dg}(\boldsymbol{F}_t)\boldsymbol{H}_t\mathrm{d}\boldsymbol{W}_t\end{aligned}\tag{7.4.49}$$

注意，当 $\mathrm{d}t$ 趋近于零时，式(7.4.49)中仅有 $O(\mathrm{d}t)$这一项有意义。因为对于无穷小的 $\mathrm{d}t$ 来说，方程右边的第二项和第三项将变成零。方程右边的第四项为 $\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(M_t)_{ij}$ $\mathrm{d}W_i^{(i)}\mathrm{d}W_i^{(j)}$，其中 $\boldsymbol{M}_t=\boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{Dg}(\boldsymbol{F}_t)\boldsymbol{H}_t$。因为$(\mathrm{d}W_t^{(i)})^2\sim\mathrm{d}t$，$\boldsymbol{W}_t$ 的分量相互独立，所以有 $\mathrm{d}W_t^{(i)}\mathrm{d}W_t^{(j)}\sim o(\mathrm{d}t)$成立，进而可将第四项近似为 $\frac{1}{2}\sum_{i=1}^{m}(\boldsymbol{M}_t)_{ii}\ \mathrm{d}t=\frac{1}{2}\mathrm{tr}(\boldsymbol{H}_t^{\mathrm{T}}\boldsymbol{Dg}(\boldsymbol{F}_t)\boldsymbol{H}_t)\mathrm{d}t$，即式(7.4.48)成立。

2）式(7.4.47)中定义的过程 $\mathrm{d}\langle F^{(i)},F^{(j)}\rangle_t$ 也称为 $F_t^{(i)}$ 和 $F_t^{(j)}$ 的二次变分过程。对于单个 Ito 过程 $\mathrm{d}F_t=F_t\mathrm{d}t+H_t\mathrm{d}W_t$ 来说，该二次变分可以简单地表示为 $\mathrm{d}\langle F,F\rangle_t=H_t^2\mathrm{d}t$。对于向量而言，因为各个分量之间没有耦合，因此 H_t 是对角矩阵，即对所有的$i\neq j$，有 $\mathrm{d}\langle F^{(i)},F^{(i)}\rangle_t=(H_{ii})^2\mathrm{d}t$和 $\mathrm{d}\langle F^{(i)},F^{(j)}\rangle_t=0$ 成立。

现在再回到求解 $\mathrm{d}\,\widehat{f(X_t)}$的问题上。首先考虑随机过程 $M_t\triangleq f(X_t)\Lambda_t(X_0^t;Z_0^t)$。可以将此过程写为 $f(X_t)\mathrm{e}^{F_t}$，其中

$$\mathrm{d}F_t=h(X_t,t)\mathrm{d}Z_t-\frac{1}{2}h^2(X_t,t)\mathrm{d}t=\frac{1}{2}h^2(X_t,t)\mathrm{d}t-h(X_t,t)\mathrm{d}V_t\tag{7.4.50}$$

同时，X_t 满足扩散方程，即

$$\mathrm{d}X_t=m(X_t,t)\mathrm{d}t+\sigma(X_t,t)\mathrm{d}W_t\tag{7.4.51}$$

又因为 V_t 和 W_t 是独立的维纳过程，所以 F_t/X_t 是形如式(7.4.45)的向量形式的 Ito 过程，其中 $m=k=2$，$\boldsymbol{G}_t$ 和 $\boldsymbol{H}_t$ 分别为

$$G_t = \begin{bmatrix} m(X_t,t) \\ \frac{1}{2}h^2(X_t,t) \end{bmatrix}, \quad H_t = \begin{bmatrix} \sigma(X_t,t) & 0 \\ 0 & -h(X_t,t) \end{bmatrix}$$

记 $M_t = g(X_t, F_t)$，其中 $g(x_1, x_2) = f(x_1 e^{x_2})$，可以对 M_t 应用向量形式的 Ito 微分规则。注意

$$\nabla g(x) = \begin{pmatrix} f'(x_1)e^{x_2} \\ f(x_1)e^{x_2} \end{pmatrix}, \quad \boldsymbol{D}g(x) = \begin{pmatrix} f''(x_1)e^{x_2} & f'(x_1)e^{x_2} \\ f'(x_1)e^{x_2} & f(x_1)e^{x_2} \end{pmatrix}$$

且有

$$d\langle X,F\rangle_t = d\langle F,X\rangle_t = 0$$
$$d\langle X,X\rangle_t = \sigma^2(X_t,t)dt$$
$$d\langle F,F\rangle_t = h^2(X_t,t)dt$$

假设 f''连续，可以得到

$$dM_t = f'(X_t)e^{F_t}dX_t + f(X_t)e^{F_t}dF_t + \frac{1}{2}f''(X_t)e^{F_t}\sigma^2(X_t,t)dt + \frac{1}{2}f(X_t)e^{F_t}h^2(X_t,t)dt \tag{7.4.52}$$

根据 $\Lambda_t = e^{F_t}$，$M_t = f(X_t)\Lambda_t$ 和 $dF_t = h(X_t, t)dZ_t - \frac{1}{2}h^2(X_t, t)dt$，式(7.4.52)可以重写为关于 X_0^t 和 Z_0^t 的表达式，即

$$d[f(X_t)\Lambda_t] = f'(X_t)\Lambda_t dX_t + f(X_t)h(X_t,t)\Lambda_t dZ_t + \frac{1}{2}f''(X_t)\sigma^2(X_t,t)\Lambda_t dt \tag{7.4.53}$$

对于函数 $\alpha(x_0^t)$，令$\overline{\alpha(x_0^t)}$表示积分$\int \alpha(x_0^t)\Lambda_t(x_0^t;Z_0^t)P_{X_0^t}(dx_0^t)$。[注意，$\overline{\alpha(x_0^t)}$是 Z_0^t 的函数。]因此，当 $N_t(Z_0^t) = \overline{f(X_t)}$且 $D_t(Z_0^t) = \overline{1}$时，在式(7.4.53)两边同时应用该线性算子，可以得到关于 $dN_t(Z_0^t)$的方程，即

$$dN_t(Z_0^t) = \overline{f'(X_t)dX_t} + \overline{[f(X_t)h(X_t,t)]}dZ_t + \frac{1}{2}\overline{f''(X_t)\sigma^2(X_t,t)}dt \tag{7.4.54}$$

第一项$\overline{f'(X_t)dX_t}$是在固定 Z_0^t(而不依赖于 Z_0^t)的条件下，$f'(X_t)\Lambda_t(X_0^t; Z_0^t)dX_t$ 中 X_0^t 的非条件期望。由于 $dX_t = m(X_t, t)dt + \sigma(X_t, t)dW_t$ 且 dW_t 与所有过去时刻无关，所以可以得到$\overline{f'(X_t)dX_t} = \overline{f'(X_t)m(X_t, t)dt}$。因此方程(7.4.54)变为

$$\begin{aligned} dN_t(Z_0^t) &= \overline{\left[f'(X_t)m(X_t,t) + \frac{1}{2}f''(X_t)\sigma^2(X_t,t)\right]}dt + \overline{f(X_t)h(X_t,t)}dZ_t \\ &= \overline{\mathcal{A}_t f(X_t)}dt + \overline{f(X_t)h(X_t,t)}dZ_t \\ &= [\overline{\mathcal{A}_t f(X_t)} + \overline{f(X_t)h(X_t,t)}h(X_t,t)]dt + \overline{f(X_t)h(X_t,t)}dV_t \end{aligned} \tag{7.4.55}$$

由于 $D_t(Z_0^t)$是 $f(x)=1$ 时 $N_t(Z_0^t)$的特例，于是有

$$dD_t(Z_0^t) = \overline{h(X_t,t)}h(X_t,t)dt + \overline{h(X_t,t)}dV_t \tag{7.4.56}$$

又由于 V_t 是维纳过程，可从式(7.4.55)和式(7.4.56)中看出，N_t/D_t 构成 $m=2$、$k=1$ 的向量 Ito 过程，其中

$$
\boldsymbol{G}_t = \begin{pmatrix} \overline{\mathcal{A}_t F(X_t)} + \overline{f(X_t)h(X_t,t)}h(X_t,t) \\ \overline{h(X_t,t)}h(X_t,t) \end{pmatrix}
$$

且

$$
\boldsymbol{H}_t = \begin{pmatrix} \overline{(X_t)h(X_t,t)} \\ \overline{h(X_t,t)} \end{pmatrix}
$$

根据式(7.4.44)可以发现，$\widehat{f(X_t)} = N_t(Z_0^t)/D_t(Z_0^t) = g(N_t,\ D_t)$，其中 $g(x_1,\ x_2) = x_1/x_2$。求偏导后，有

$$
\nabla g(\boldsymbol{x}) = \begin{pmatrix} 1/x_2 \\ -x_1/x_2^2 \end{pmatrix} \quad 且 \quad \boldsymbol{D}\boldsymbol{g}(\boldsymbol{x}) = \begin{pmatrix} 0 & -1/x_2^2 \\ -1/x_2^2 & 2x_1/x_2^3 \end{pmatrix}
$$

同样有

$$
\mathrm{d}\langle N,D\rangle_t = \mathrm{d}\langle D,N\rangle_t = \overline{[f(X_t)h(X_t,t)][h(X_t,t)]}\mathrm{d}t
$$

$$
\mathrm{d}\langle D,N\rangle_t = \overline{[f(X_t)h(X_t,t)]^2}\,\mathrm{d}t
$$

和

$$
\mathrm{d}\langle D,D\rangle_t = \overline{[h(X_t,t)]^2}\,\mathrm{d}t
$$

结合上述公式并应用向量形式的 Ito 公式[⊖]后，可以得到

$$
\mathrm{d}\,\widehat{f(X_t)} = \frac{1}{D_t}\mathrm{d}N_t - \frac{N_t}{D_t^2}\mathrm{d}D_t - \frac{1}{D_t^2}[\overline{f(X_t)h(X_t,t)}][\overline{h(X_t,t)}]\mathrm{d}t + \frac{N_t}{D_t^3}[\overline{h(X_t,t)}]^2\,\mathrm{d}t
\tag{7.4.57}
$$

根据式(7.4.55)和式(7.4.56)，有

$$
\mathrm{d}\,\widehat{f(X_t)} = \left[\frac{\overline{\mathcal{A}_t F(X_t)}}{D_t} - \frac{[\overline{f(X_t)h(X_t,t)}]}{D_t}\,\frac{\overline{h(X_t,t)}}{D_t} + \frac{N_t}{D_t}\left[\frac{[\overline{h(X_t,t)}]}{D_t}\right]^2\right]\mathrm{d}t
+ \left[\frac{[\overline{f(X_t)h(X_t,t)}]}{D_t} - \frac{N_t}{D_t}\,\frac{[\overline{h(X_t,t)}]}{D_t}\right]\mathrm{d}Z_t
\tag{7.4.58}
$$

注意，对于任何具有 $g(X_t)$形式的随机变量而言，有$\widehat{g(X_t)} = \overline{g(X_t)}/D_t$ 成立。因此，结合 $\widehat{f(X_t)} = N_t/D_t$ 和更新项的定义式($\mathrm{d}I_t = \mathrm{d}Z_t - \hat{h}\ \mathrm{d}t$)后，式(7.4.58)退化为

$$
\mathrm{d}\,\widehat{f(X_t)} = \widehat{\mathcal{A}_t F(X_t)}\mathrm{d}t + [\widehat{f(X_t)h(X_t,t)} - \widehat{f(X_t)}\ \widehat{h(X_t,t)}]\mathrm{d}I_t
\tag{7.4.59}
$$

这就是 $R_t \equiv 1$ 时的滤波方程(7.4.28)。令 $\Lambda_0 \equiv 1$ 后，初始条件$\widehat{f(X_0)}$可以根据式(7.4.44)求解。

备注

1) 未归一化的条件密度。根据式(7.4.55)，$\overline{f(X_t)}$满足演化方程

$$
\mathrm{d}\,\overline{f(X_t)} = \overline{\mathcal{A}_t F(X_t)}\mathrm{d}t + \overline{f(X_t)h(X_t,t)}\mathrm{d}Z_t
\tag{7.4.60}
$$

这相比$\widehat{f(X_t)}$的变化方程稍简单。既然根据$\widehat{f(X_t)}$的方程可得到关于 $q_{X_t}(x)$的演化方程，而根据式(7.4.60)可以得到与 $q_{X_t}(x)$有关的“密度函数”，这也就意味着能找到比 $q_{X_t}(x)$更加简单的方程。特别地，假设定义

⊖ 注意，∇g 和 $\boldsymbol{D}\boldsymbol{g}$ 在 $x_2=0$ 处是不连续的。然而，D_t 以概率 1 为正(相应的原因将在下面讨论)，所以该不连续点完全可以忽略。

$$\rho_{X_t}(x) = \overline{\delta(x - X_t)} \equiv \int \delta(x - x_t) \Lambda_t(x_0^t ; Z_0^t) P_{X_0^t}(\mathrm{d}x_0^t) \tag{7.4.61}$$

其中 δ 是狄拉克冲激函数。[注意，$\rho_{X_t}(x)$是 Z_0^t 函数。]进而有

$$q_{X_t}(x) = \widehat{\delta(x - X_t)} = \frac{\overline{\delta(x - X_t)}}{N_t} = \frac{\rho_{X_t}(x)}{N_t}$$

而且易知

$$\overline{f(X_t)} = \int_{-\infty}^{\infty} f(x) \rho_{X_t}(x) \mathrm{d}x \tag{7.4.62}$$

因此 $N_t \triangleq \overline{1} = \int_{-\infty}^{\infty} \rho_{X_t}(x) \mathrm{d}x$ 且

$$q_{X_t}(x) = \frac{\rho_{X_t}(x)}{\int_{-\infty}^{\infty} \rho_{X_t}(x) \mathrm{d}x} \tag{7.4.63}$$

根据式(7.4.63)可知，对于每个 t 而言，ρ_{X_t} 是给定 Z_0^t 条件下 X_t 的非归一化条件密度，它和 q_{X_t} 有着相同的形式，但积分可能不等于 1。由于 q_{X_t} 可以从 ρ_{X_t} 中得到，就计算而言，二者等价。

结合式(7.4.60)和式(7.4.62)，有

$$\begin{aligned} \mathrm{d}\int_{-\infty}^{\infty} f(y) \rho_{X_t}(y) \mathrm{d}y &= \left[\int_{-\infty}^{\infty} \mathcal{A}_t f(y) \rho_{X_t}(y) \mathrm{d}y\right] \mathrm{d}t + \left[\int_{-\infty}^{\infty} f(y) h(y,t) \rho_{X_t}(y) \mathrm{d}y\right] \mathrm{d}Z_t \\ &= \int_{-\infty}^{\infty} f(y)\, \mathcal{A}_t^* \rho_{X_t}(y) \mathrm{d}t + \left[\int_{-\infty}^{\infty} f(y) h(y,t) \rho_{X_t}(y) \mathrm{d}y\right] \mathrm{d}Z_t \end{aligned} \tag{7.4.64}$$

令 $f(y)$趋近 $\delta(x-y)$后，就可推导出 ZaKai 方程，即

$$\mathrm{d}\rho_{X_t}(x) = \mathcal{A}_t^* \rho_{X_t}(x) \mathrm{d}t + h(x,t) \rho_{X_t}(x) \mathrm{d}Z_t \tag{7.4.65}$$

作为对比，考虑 $R_t \equiv 1$ 时 $q_{X_t}(x)$的演化过程，即

$$\mathrm{d}q_{X_t}(x) = \mathcal{A}_t^* q_{X_t}(x) \mathrm{d}t + q_{X_t}(x)[h(x,t) - \widehat{h(X_t,t)}][\mathrm{d}Z_t - \widehat{h(X_t,t)} \mathrm{d}t] \tag{7.4.66}$$

比较 ρ_{X_t} 和 q_{X_t} 的演化方程，可以看出前者更简洁。特别地，由于 $\widehat{h(X_t,t)} = \int_{-\infty}^{\infty} h(x,t) \cdot q_{X_t}(x) \mathrm{d}x$且存在 $q\,\hat{h}\,\mathrm{d}Z$、$qh\,\hat{h}\,\mathrm{d}t$ 以及$(q\,\hat{h})^2 \mathrm{d}t$ 项，所以式(7.4.66)是关于 q_{X_t} 的非线性方程。与此相反，式(7.4.65)关于 ρ_{X_t} 线性，因此更容易分析。[注意，式(7.4.65)并不是线性微分方程，因为方程的右边存在 ρ_{X_t} 和 $\mathrm{d}Z_t$ 的乘积。不过，该方程却是双线性的，即对于 ρ_{X_t} 和 $\mathrm{d}Z_t$ 来说，只要其中一个固定，则另一个是线性的。]此外，式(7.4.65)更加容易计算。在给定 x 时，对 ρ_{X_t} 进行数值计算只需要知道 x 下的 ρ_{X_t}、ρ'_{X_t} 和 ρ''_{X_t}。这些量都是可以根据 ρ_{X_t} 在给定 x 的邻域内的分布计算出来的局部特性。相反，计算给定 x 时 q_{X_t} 的数值传递则需要求解积分 $\int_{-\infty}^{\infty} h(x,t) q_{X_t}(x) \mathrm{d}x$，它是由 q_{X_t} 在所有 x 下计算出来的全局量。

事实上，非归一化密度 ρ_{X_t} 在非线性滤波中有着更重要的意义，即在非线性滤波中，可以通过 ρ_{X_t} 的演化方程来求解有限维的最优滤波器。关于此问题的讨论将会在下面给出。

2) 检测由扩散方程产生的信号的估计-相关器公式。信号检测问题可用假设检验对描述为

$$\begin{aligned} H_0 &: \mathrm{d}Z_t = \mathrm{d}V_t, \quad 0 \leqslant t \leqslant 1 \\ H_1 &: \mathrm{d}Z_t = S_t \mathrm{d}t + \mathrm{d}V_t, \quad 0 \leqslant t \leqslant 1 \end{aligned} \tag{7.4.67}$$

其中信号 S_t 是扩散过程$\{X_t;\ 0\leqslant t\leqslant 1\}$的函数，即 $S_t=h(X_t,\ t)$，$\{V_t;\ 0\leqslant t\leqslant 1\}$是独立于$\{X_t;\ 0\leqslant t\leqslant 1\}$的维纳过程且 $\int_0^1 S_t^2\,\mathrm{d}t<\infty$ 依概率 1 成立。

假设观测持续到某个时间 $t\in(0,\ 1)$。令 $X_0^t=x_0^t$，式(7.4.67)的似然比由 Cameron-Martin 公式给出

$$\exp\left\{\int_0^t h(x_u,u)\mathrm{d}Z_u-\frac{1}{2}\int_0^t h^2(x_u,u)\mathrm{d}u\right\}$$

这就是 $\Lambda_t(x_0^t;\ Z_0^t)$。因此，持续到时间 t 的观测的似然比 L_t 可以通过对 $\Lambda_t(x_0^t;\ Z_0^t)$在 X_0^t 的分布上求取平均得到，即

$$L_t=\int\Lambda_t(x_0^t;Z_0^t)PX_0^t(\mathrm{d}x_0^t)=\overline{1}=D_t(Z_0^t)$$

其中 D_t 是式(7.4.44)中的分母项。根据式(7.4.56)，L_t 满足随机微分方程，即

$$\mathrm{d}L_t=\overline{h(X_t,t)}\mathrm{d}Z_t=\overline{S}_t\mathrm{d}Z_t=\hat{S}_tL_t\mathrm{d}Z_t \tag{7.4.68}$$

此处利用了$\hat{S}_t=\overline{S}_t/D_t$。

再考虑 $\log L_1$，有

$$\mathrm{d}L_t=\hat{S}_tL_tS_t\mathrm{d}t+\hat{S}_tL_t\mathrm{d}V_t$$

并且利用 $g(x)=\log(x)$条件下的 Ito 公式，有

$$\mathrm{d}\log L_t=\frac{1}{L_t}\mathrm{d}L_t-\frac{1}{2}\left(\frac{1}{L_t^2}\right)(\hat{S}_tL_t)^2\mathrm{d}t=\hat{S}_t\mathrm{d}Z_t-\frac{1}{2}(\hat{S}_t)^2\mathrm{d}t \tag{7.4.69}$$

式(7.4.69)的积分形式为

$$\log L_t=\log L_0+\int_0^t\hat{S}_u\mathrm{d}Z_u-\frac{1}{2}\int_0^t[\hat{S}_u]^2\mathrm{d}u \tag{7.4.70}$$

又因为 $L_0=1$，$L_1=\mathrm{d}P_1/\mathrm{d}P_0(Z_0^t)$，那么式(7.4.70)意味着

$$\frac{\mathrm{d}P_1}{\mathrm{d}P_0}(Z_0^1)=\exp\left\{\int_0^1\hat{S}_t\mathrm{d}Z_t-\frac{1}{2}\int_0^1[\hat{S}_t]^2\mathrm{d}t\right\} \tag{7.4.71}$$

这就是在 6.4 节中给出的似然比估计器-相关器公式。

注意，在推导式(7.4.71)时，并没有直接利用新息过程 $\mathrm{d}I_t=\mathrm{d}Z_t-\hat{S}_t\mathrm{d}t$ 的任何性质。正如 6.4 节所述，新息过程所具有的特性在本质上遵循似然比公式(7.4.71)。此外，尽管式(7.4.71)的有效性最早在扩散模型上得到了验证[参见 Duncan(1968)或者 Stratonovich 和 Sosulin(1965)]，但它在一般意义上也是成立的。

3) 非线性滤波中的有限维度。正如之前已经提到的，条件密度(归一化的或非归一化的)的演化方程具有无限维递归特性。在非线性模型中求解有限维的最优滤波器在本质上需要求解一组有限维的变量，这些变量能够独立于其他量传播，并能由此计算出条件概率密度的充分统计量。事实上，高斯 Kalman-Bucy 模型恰好说明了这一点，在该模型中，所有的条件概率密度是高斯的，因而仅由其均值和方差所决定。正如式(7.4.31)和式(7.4.42)所示，对于 Kalman-Bucy 模型，均值和方差的更新/递推过程与其他量无关。

在非线性滤波模型中可，以利用非归一化的密度方程 $\mathrm{d}\rho=\mathcal{A}_s^*\rho\mathrm{d}t+h\rho\mathrm{d}Z_t$ 检验是否存在有限维充分统计量。研究结果表明：由于该方程是双线性的，所以利用向量域有关的 Lie 代数的有限维度可以回答该方程有限维可实现性的问题。此处不再详述这一理论，只

给出两个有趣的情况。其中一个是立方传感器问题，即

$$dX_t = dW_t$$
$$dZ_t = X_t^3 dt + dV_t$$

尽管它的表达式很简单，但在本质上具有无限维特性[参见 Hazewinkel 和 Marcus (1982)]；另一类是所谓的 Beneš 类问题，即

$$dX_t = m(X_t)dt + dW_t$$
$$dZ_t = X_t dt + dV_t$$

其中，对于常数 a、b 和 $c(a \geqslant -1)$，有 $m'(x)+m^2(x)=ax^2+bx+c$ 成立。研究表明该问题具有十维充分统计量[参见 Benes(1981)]。

读者若是对非线性滤波的细节感兴趣，可以参考 Brockett 和 clark(1980)或者 Marcus (1984)的文献。

7.4.3 最优非线性滤波器的近似方法

只有在某些特殊环境下，最优非线性滤波方程才能以有限维实现，因此将研究重点转移到如何对这些方程进行有限维逼近。

例如，考虑直接状态估计器的演化公式(7.4.30)

$$\begin{aligned} d\hat{X}_t &= \widehat{m(X_t,t)}dt + [\widehat{X_t h(X_t,t)} - \hat{X}_t \widehat{h(X_t,t)}]R_t^{-1}dI_t \\ &= \widehat{m(X_t,t)}dt + [\widehat{X_t h(X_t,t)} - \hat{X}_t \widehat{h(X_t,t)}]R_t^{-1}dZ_t \\ &\quad - [\widehat{X_t h(X_t,t)} - \hat{X}_t \widehat{h(X_t,t)}]R_t^{-1}\widehat{h(X_t,t)}dt \end{aligned}$$

求解该方程的难题在于方程右边存在$\hat{m}$、$\widehat{Xh}$和$\hat{h}$项。但是，如果假设估计误差 $X_t-\hat{X}_t$ 在大部分时间内都较小且 $m(x, t)$、$h(x, t)$都是 x 的光滑函数，$m(x, t)$和 $h(x, t)$可近似为

$$m(X_t,t) \approx m(\hat{X}_t,t) + (X_t - \hat{X}_t)m'(\hat{X}_t,t) \tag{7.4.72}$$

和

$$h(X_t,t) \approx h(\hat{X}_t,t) + (X_t - \hat{X}_t)h'(\hat{X}_t,t) \tag{7.4.73}$$

其中，$m'(x, t)=\partial m(x, t)/\partial x$，$h'(x, t)=\partial h(x, t)/\partial x$。现在，考虑式(7.4.72)和式(7.4.73)中的条件期望，因为 $E\{X_t-\hat{X}_t|Z_0^t\}=0$ 后，所以有

$$\widehat{m(X_t,t)} \approx m(\hat{X}_t,t) \tag{7.4.74}$$

和

$$\widehat{h(X_t,t)} \approx h(\hat{X}_t,t) \tag{7.4.75}$$

类似地，$Xh-\hat{X}\hat{h}$也可近似为

$$\begin{aligned} \widehat{X_t h(X_t,t)} - \hat{X}_t \widehat{h(X_t,t)} &= \widehat{(X_t - \hat{X}_t)h(X_t,t)} \\ &\approx \widehat{(X_t - \hat{X}_t)^2}h'(\hat{X}_t,t) = P_t h'(\hat{X}_t,t) \end{aligned} \tag{7.4.76}$$

其中 P_t 是给定 Z_0^t 后 X_t 的条件方差。将式(7.4.74)代入式(7.4.76)中，得到条件均值演

化的近似表达式为

$$\mathrm{d}\,\hat{X}_t \approx m(\hat{X}_t,t)\mathrm{d}t + P_t h'(\hat{X}_t,t)R_t^{-1} \times [\mathrm{d}Z_t - h(\hat{X}_t,t)\mathrm{d}t] \tag{7.4.77}$$

注意，式(7.4.77)的右边只涉及 $\mathrm{d}Z_t$、$\hat{X}_t$ 和 P_t。因此，如果可以计算出 P_t，就可以实现近似滤波器。

下面考虑方差的传递方程(7.4.38)，即

$$\begin{aligned}\mathrm{d}P_t =& [2\,\widehat{(X_t-\hat{X}_t)m(\hat{X}_t,t)} + \widehat{\sigma^2(X_t,t)} - [\widehat{X_t h(X_t,t)} - \hat{X}_t\,\widehat{h(X_t,t)}]^2 R_t^{-1}]\mathrm{d}t \\ &+ [\widehat{(X_t-\hat{X}_t)^2 h(X_t,t)} - P_t\,\widehat{h(X_t,t)}]R_t^{-1}]\mathrm{d}I_t\end{aligned} \tag{7.4.78}$$

如果采用与前面相同的方法来逼近该方程，并对 $\sigma^2(x,t)$进行一阶泰勒级数近似后，有

$$\begin{aligned}\mathrm{d}P_t \approx& [2P_t m'(\hat{X}_t,t) + \sigma^2(\hat{X}_t,t) - P_t^2[h'(\hat{X}_t,t)]^2 R_t^{-1}]\mathrm{d}t \\ &+ [\widehat{(X_t-\hat{X}_t)^3} h'(\hat{X}_t,t)R_t^{-1}[\mathrm{d}Z_t - h(\hat{X}_t,t)]\mathrm{d}t]\end{aligned} \tag{7.4.79}$$

然而，相比式(7.4.77)而言，式(7.4.79)的右边除包含 $\mathrm{d}Z_t$、$\hat{X}_t$ 和 P_t 之外，还包含了条件三阶中心距$\widehat{(X_t-\hat{X}_t)^3}$。如果重复上述逼近过程，就会得到依赖于条件四阶中心距的近似方程。因此，即使逼近后，仍是无限维方程。不过，如果进一步假设，误差$X_t-\hat{X}_t$关于零均值对称分布(例如高斯误差)，则有$\widehat{(X_t-\hat{X}_t)^3}=0$ 成立，此时式(7.4.77)和式(7.4.79)就会变成一组封闭的方程，即

$$\mathrm{d}\,\hat{X}_t = m(\hat{X}_t,t)\mathrm{d}t + P_t h'(X_t,t)R_t^{-1} \times [\mathrm{d}Z_t - h(\hat{X}_t,t)]\mathrm{d}t \tag{7.4.80}$$

和

$$\dot{P}_t = 2P_t m'(\hat{X}_t,t) + \sigma^2(\hat{X}_t,t) - P_t^2[h'(\hat{X}_t,t)]^2 R_t^{-1} \tag{7.4.81}$$

由式(7.4.80)和式(7.4.81)所描述的滤波器称为扩展卡尔曼滤波器，常用于逼近最优非线性滤波器。图 7.4.1 中给出该滤波器的实现框图。与卡尔曼滤波器不同的是，在扩展卡尔曼滤波器中需要把状态估计输入到计算增益的滤波器中。下面，通过一个例子说明这些方程的应用。

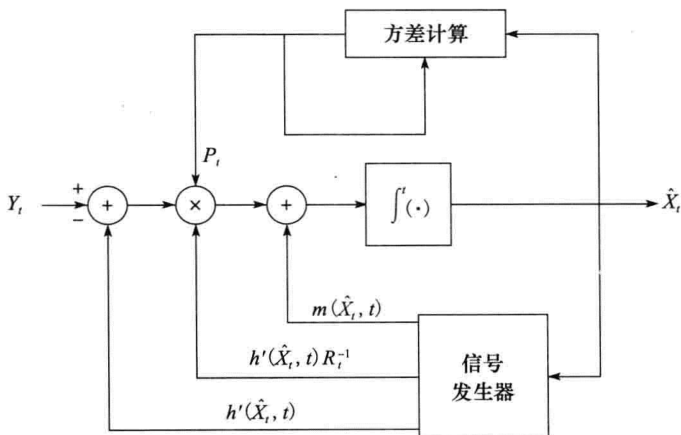


图 7.4.1　扩展卡尔曼滤波器

例 7.4.3　相位跟踪

相位跟踪问题是实际应用中经常出现的典型非线性估计问题，在该问题中，需要利用噪声环境下的正弦信号观测值来估计正弦信号的相位。特别地，考虑如下观测模型

$$\mathrm{d}Z_t = C\sin(\omega_0 t + X_t)\mathrm{d}t + \left[\frac{N_0}{2}\right]^{1/2}\mathrm{d}V_t \tag{7.4.82}$$

其中，C 和 ω_0 分别是已知的常数振幅和频率，噪声谱密度为 $N_0/2$。该模型对应 $h(x, t)=A\sin(\omega_0 t+x)$，此时扩展卡尔曼滤波方程变为

$$\mathrm{d}\,\hat{X}_t = m(\hat{X}_t, t)\mathrm{d}t + (2CP_t/N_0)\cos(\omega_0 t + \hat{X}_t) \times [\mathrm{d}Z_t - C\sin(\omega_0 t + \hat{X}_t)\mathrm{d}t] \tag{7.4.83}$$

和

$$\dot{P}_t = 2P_t m'(\hat{X}_t, t) + \sigma^2(\hat{X}_t, t) - 2C^2 P_t^2 \cos^2(\omega_0 t + \hat{X}_t)/N_0 \tag{7.4.84}$$

通常可将相位扩散项 X_t 视为由一阶时不变线性模型产生，即

$$\mathrm{d}X_t = -aX_t\mathrm{d}t + \mathrm{d}W_t \tag{7.4.85}$$

其中 $a>0$。此时，扩展卡尔曼滤波方程简化为

$$\mathrm{d}\,\hat{X}_t = -a\,\hat{X}_t + (2CP_t/N_0)\cos(\omega_0 t + \hat{X}_t) \times [\mathrm{d}Z_t - C\sin(\omega_0 t + \hat{X}_t)\mathrm{d}t] \tag{7.4.86}$$

和

$$\dot{P}_t = -2aP_t + 1 - 2C^2 P_t^2 \cos^2(\omega_0 t + \hat{X}_t)/N_0 \tag{7.4.87}$$

式(7.4.86)本质上是一个受非线性反馈$(2CP_t/N_0)\cos(\omega_0 t+\hat{X}_t)[\mathrm{d}Z_t-\sin(\omega_0 t+\hat{X}_t)\mathrm{d}t]$激励的低通滤波器($\mathrm{d}\,\hat{X}_t=-a\,\hat{X}_t\mathrm{d}t$)，其 3dB 带宽为 $2a$。典型相位带宽(也为 $2a$)远小于载波频率 ω_0。此时高频信号 $\cos(\omega_0 t+\hat{X}_t)\sin(\omega_0 t+\hat{X}_t)=(1/2)\sin(2\omega_0 t+2\,\hat{X}_t)$无法通过估计滤波器。因此，该估计器可近似为

$$\mathrm{d}\,\hat{X}_t = -a\,\hat{X}_t + (2CP_t/N_0)\cos(\omega_0 t + \hat{X}_t)\mathrm{d}Z_t \tag{7.4.88}$$

同理，式(7.4.87)中的 $2C^2P_t^2\cos^2(\omega_0 t+\hat{X}_t)/N_0$ 项等于 $C^2P_t^2[1+\cos(2\omega_0 t+2\,\hat{X}_t)]/N_0$。如果 $a\ll\omega_0$，那么该式中的第二项无法通过滤波器，进而可将方差方程简化为

$$\dot{P}_t = -2aP_t + 1 - C^2 P_t^2/N_0 \tag{7.4.89}$$

事实上，用 Cx 代替 $h(x, t)$后，就能得到 Kalman-Bucy 滤波模型中的 Riccati 方程。假设方差到达稳态值 P_∞，则相位跟踪器的稳态滤波器方程为

$$\mathrm{d}\,\hat{X}_t = -a\,\hat{X}_t + K_\infty\cos(\omega_0 t + \hat{X}_t)\mathrm{d}Z_t \tag{7.4.90}$$

其中 $K_\infty=2CP_\infty/N_0$。

图 7.4.2 给出了式(7.4.90)所描述的非线性滤波器(通常 $\mathrm{d}Z_t=Y_t\mathrm{d}t$)。该滤波器即为锁相环，在实际中应用非常广泛。注意，锁相环的稳态性能 P_∞ 是下列方程的正根

$$0 = -2aP_\infty + 1 - C^2 P_\infty^2/N_0$$

即

$$P_\infty = \frac{aN_0}{C^2}\left(\left[1+\frac{C^2}{aN_0}\right]^{1/2} - 1\right)$$

C^2/aN_0 可视为锁相环中的信噪比(SNR)，因为 $C\sin(\omega_0 t+X_t)$的平均功率正比于 C^2 和谱密度为 $N_0/2$ 的白噪声通过 3dB 带宽为 $2a$ 的滤波器后的平均输出功率为 aN_0。当 C^2/aN_0

增大时，有 $P_\infty \sim [aN_0/C^2]^{1/2}$，即稳态跟踪精度随着信噪比(SNR)平方根的倒数降低而降低。注意，在高信噪比(SNR)情况下，扩展卡尔曼滤波器的性能会最接近最小均方误差非线性滤波器的性能，这是因为 $X_t-\hat{X}_t$ 非常小。

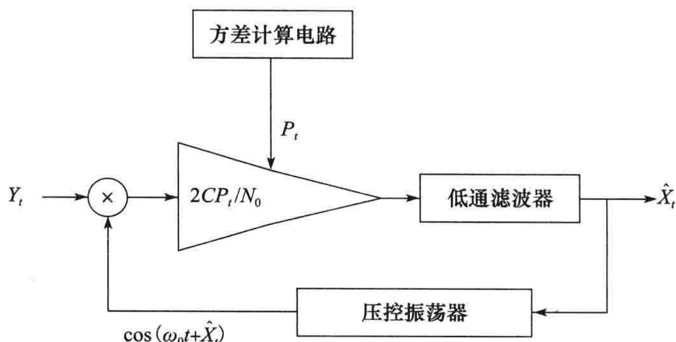


图 7.4.2 锁相环：最优相位跟踪器的逼近

读者若是对锁相环的内容及其性能感兴趣，可以参考 Viterbi 的书(1968)。 ◀

在很多应用中，扩展卡尔曼滤波器能够有效逼近最优非线性滤波器。但是，如果平均估计误差不够小，那么该滤波器对最优非线性滤波器的近似将很差。此时，需要利用高阶泰勒级数展开来更好地逼近非线性过程。

二阶滤波器是对卡尔曼滤波器的扩展，它利用二阶来近似非线性。特别地，将 m 展开为

$$m(X_t,t)\approx m(\hat{X}_t,t)+(X_t-\hat{X}_t)m'(\hat{X}_t,t)+\frac{1}{2}(X_t-\hat{X}_t)^2m''(\hat{X}_t,t)$$

对 h 进行类似的操作，滤波器方程中估计量的二阶近似为

$$\widehat{m(X_t,t)}\approx m(\hat{X}_t,t)+\frac{1}{2}P_tm''(\hat{X}_t,t)$$

$$\widehat{h(X_t,t)}\approx h(\hat{X}_t,t)+\frac{1}{2}P_th''(\hat{X}_t,t)$$

和

$$\widehat{(X_t-\hat{X}_t)h(X_t,t)}\approx\left[P_th'(\hat{X}_t,t)+\frac{1}{2}\widehat{(X_t-\hat{X}_t)^3}h''(\hat{X}_t,t)\right]$$

再次假设 $\widehat{(X_t-\hat{X}_t)^3}=0$ 后，$\hat{X}_t$ 的二阶演化方程变为

$$\begin{aligned}\mathrm{d}\,\hat{X}_t\approx&\left[m(\hat{X}_t,t)+\frac{1}{2}P_tm''(\hat{X}_t,t)\right]\mathrm{d}t\\&+P_th'(\hat{X}_t,t)R_t^{-1}\left[\mathrm{d}Z_t-\left[h(\hat{X}_t,t)-\frac{1}{2}P_th''(\hat{X}_t,t)\right]\mathrm{d}t\right]\end{aligned}\tag{7.4.91}$$

求解上述问题还需要已知 P_t 的演化方程。应用上述步骤[包括令 $\widehat{(X_t-\hat{X}_t)^3}=0$]且利用

$$\begin{aligned}\widehat{(X_t-\hat{X}_t)^2h(X_t,t)}-P_t\widehat{h(X_t,t)}&=\widehat{(X_t-\hat{X}_t)^2[h(X_t,t)-\widehat{h(X_t,t)}]}\\&\approx\frac{1}{2}\widehat{(X_t-\hat{X}_t)^4}h'(\hat{X}_t,t)\end{aligned}$$

得到 P_t 的演化方程为

$$\begin{aligned} \mathrm{d}P_t \approx & \{2P_t m'(\hat{X}_t,t)+\sigma^2(\hat{X}_t,t)+P_t[\sigma'(\hat{X}_t,t)]^2+P_t\sigma(\hat{X}_t,t)\sigma''(\hat{X}_t,t) \\ & -P_t^2[h'(\hat{X}_t,t)]^2 R_t^{-1}\}\mathrm{d}t+\frac{1}{2}\widehat{(X_t-\hat{X}_t)^4}h''(\hat{X}_t,t)R_t^{-1} \\ & \times\left[\mathrm{d}Z_t-(h(\hat{X}_t,t)+\frac{1}{2}P_t h''(\hat{X}_t,t))\mathrm{d}t\right] \end{aligned} \tag{7.4.92}$$

现在有必要讨论条件四阶中心矩$\widehat{(X_t-\hat{X}_t)^4}$的求解方法。如果尝试推导它的二阶近似演化方程，那么将会涉及条件五阶和六阶中心矩。虽然对称误差假设可避免求解五阶矩，但是计算六阶矩将会涉及八阶矩等。因此必须引入一些其他的假设来消除高阶矩之间的耦合。注意，如果误差$(X_t-\hat{X}_t)$是高斯的，那么高阶矩之间的耦合将会被破坏，因为高斯随机变量的四阶中心矩和二阶中心矩的计算关系为

$$\widehat{(X_t-\hat{X}_t)^4}=3P_t^2 \tag{7.4.93}$$

在缺乏更符合实际的假设时，通常可假设式(7.4.93)近似成立，那么 P_t 的方程简化为

$$\begin{aligned} \mathrm{d}P_t \approx & \{P_t[2m'(\hat{X}_t,t)+\sigma(\hat{X}_t,t)\sigma''(\hat{X}_t,t)+(\sigma'(\hat{X}_t,t))^2] \\ & +\sigma^2(\hat{X}_t,t)-P_t^2[h'(\hat{X}_t,t)]^2 R_t^{-1}\}\mathrm{d}t+\frac{3}{2}P_t^2 h''(\hat{X}_t,t)R_t^{-1} \\ & \times[\mathrm{d}Z_t-(h(\hat{X}_t,t)+\frac{1}{2}P_t h''(\hat{X}_t,t))\mathrm{d}t] \end{aligned} \tag{7.4.94}$$

在$\widehat{(X_t-\hat{X}_t)^3}\approx 0$和$\widehat{(X_t-\hat{X}_t)^4}\approx 3P_t^2$假设条件下，式(7.4.92)和式(7.4.94)是一组对非线性滤波方程二阶逼近的封闭方程。图 7.4.3 为该滤波器的示意图。在计算增益时，不仅需要向滤波器中输入状态估计量$\hat{X}_t$，而且还需要利用观测量 Y_t。

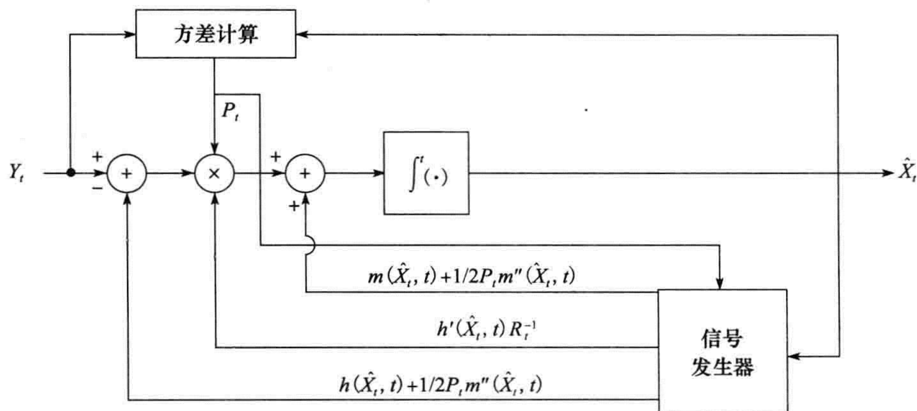


图 7.4.3 二阶滤波器

作为二阶滤波器的应用实例，再次考虑例 7.4.3 中的相位跟踪问题，其相位模型 $\mathrm{d}X_t=-aX_t+\mathrm{d}W_t$。假设 $0<a\ll\omega_0$，则滤波方程和之前的近似形式保持一致，即

$$\mathrm{d}\hat{X}_t\approx -a\hat{X}_t\mathrm{d}t+(2CP_t/N_0)\cos(\omega_0 t+\hat{X}_t)\mathrm{d}Z_t \tag{7.4.95}$$

不过，方差方程变为

$$dP_t \approx \left[2aP_t + 1 - \frac{C^2 P_t^2}{2N_0}\left(5 + \frac{3}{2}P_t\right)\right]dt - \frac{3P_t^2}{N_0}C\sin(\omega_0 t + \hat{X}_t)dZ_t \qquad (7.4.96)$$

显然，相位估计器的二阶形式和扩展卡尔曼滤波器一样，唯一的不同在于二阶滤波器中的增益计算依赖于数据。

在结束如何对非线性滤波器进行有效逼近的讨论之前，需要注意：在某些情况下，线性 Kalman-Bucy 滤波器可用来逼近最优非线性滤波器。为了证明这一点，假设某个确知函数 $x_t^{(o)}(t\geqslant 0)$满足非线性微分方程

$$dx_t^{(o)} = m(x_t^{(o)}, t)dt, \quad t \geqslant 0 \qquad (7.4.97)$$

例如，$\{x_t^{(o)};\ t\geqslant 0\}$可表示某个交通工具(如飞机)在一维方向上的运动或轨迹，而函数 $m(x_t^{(o)},\ t)$可代表交通工具的动力学以及作用在上面的确定性控制力的影响。假设真实的轨迹受随机扰动影响(如飞机扰动)而偏离名义上的轨迹$\{x_t^{(o)};\ t\geqslant 0\}$。利用扩散过程可将其建模为

$$dX_t = m(X_t, t)dt + \sigma_t dW_t \qquad (7.4.98)$$

其中 $\sigma_t dW$ 表示随机扰动。

假设含噪的观测过程为

$$dZ_t = h(X_t, t)dt + R_t^{1/2}dV_t \qquad (7.4.99)$$

如前所述，假设$\{W_t\}$和$\{V_t\}$是独立的标准维纳过程。根据 Z_0^t估计 X_t 等价于利用$\widetilde{Z}_0^t$ 估计 δX_t，其中 δX_t 表示 X_t 相对名义上的轨迹的偏差，即

$$\delta X_t \triangleq X_t - x_t^{(o)}, \quad t \geqslant 0$$

而$\widetilde{Z}_0^t$ 表示观测量与 $h(x_t^{(o)},\ t)$之间的偏差，即

$$d\widetilde{Z}_t = dZ_t - h(x_t^{(o)}, t)dt$$

结合$\hat{X}_t$ 的演化方程并利用 $x_t^{(o)}$ 表示确知轨迹，很容易得到

$$d(\widehat{\delta X_t}) = [\widehat{m(X_t, t)} - m(x_t^{(o)}, t)]dt + \mathrm{cov}(\delta X_t, h(X_t, t) - h(x_t^{(o)}, t) \mid Z_0^t)R_t^{-1}$$
$$\times [d\widetilde{Z}_t - [h(X_t, t) - h(x_t^{(o)}, t)]dt] \qquad (7.4.100)$$

假设实际轨迹与名义轨迹之间的偏差很小，则有下面的表达式近似成立，即

$$m(X_t, t) - m(x_t^{(o)}, t) \approx m'(x_t^{(o)}, t)\delta X_t$$

和

$$h(X_t, t) - h(x_t^{(o)}, t) \approx h'(x_t^{(o)}, t)\delta X_t$$

进而得到近似估计器方程为

$$d(\widehat{\delta X_t}) \approx m'(x_t^{(o)}, t)\widehat{\delta X_t} + P_t h'(x_t^{(o)}, t)R_t^{-1} \times [d\widetilde{Z}_t - h'(x_t^{(o)}, t)\widehat{\delta X_t}dt] \qquad (7.4.101)$$

注意，求解式(7.4.101)只需要知道条件方差 $P_t = \mathrm{var}(X_t \mid Z_0^t) \equiv \mathrm{var}(\delta X_t \mid Z_0^t)$和观测值 $d\widetilde{Z}_t$。进一步考虑 P_t 的演化方程，将上述近似代入式(7.4.38)后，方差方程可近似为

$$dP_t \approx [2P_t m'(x_t^{(o)}, t) + \sigma_t^2 - [P_t h'(x_t^{(o)}, t)]^2 R_t^{-1}]dt$$
$$+ \widehat{(\delta X_t - \widehat{\delta X_t})^3} R_t^{-1}[d\widetilde{Z}_t - h'(x_t^{(o)}, t)\widehat{\delta X_t}dt] \qquad (7.4.102)$$

再一次假设误差关于均值对称分布，则有$\widehat{(\delta X_t - \delta\hat{X}_t)^3} = 0$ 成立，式(7.4.102)变为

$$\dot{P}_t \approx 2P_t m'(x_t^{(o)},t) + \sigma_t^2 - [P_t h'(x_t^{(o)},t)]^2 R_t^{-1} \tag{7.4.103}$$

注意，只要 $A_t = m'(x_t^{(o)},\ t)$、$B_t = \delta_t$、$C_t = h'(x_t^{(o)},\ t)$ 且 R_t 如前所述，式(7.4.101)和式(7.4.103)就是在 Kalman-Bucy 模型中的最佳滤波方程。因此，对于该模型而言，上述近似滤波器就是最优滤波器

$$\begin{aligned} \mathrm{d}(\delta X_t) &= m'(x_t^{(o)},t)(\delta X_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t \\ \mathrm{d}\,\widetilde{Z}_t &= h'(x_t^{(o)},t)(\delta X_t)\mathrm{d}t + R_t^{1/2}\mathrm{d}V_t \end{aligned} \tag{7.4.104}$$

此方程是式(7.4.98)和式(7.4.99)在名义轨迹 $x_t^{(o)}$ 上的简单线性化。校准器通过估计状态变量来将实际过程控制在名义状态轨迹上。因此，在实际的线性状态空间模型中，经常采用上述线性化操作。

7.5 习题

1. 假设信号 $Y_t = N_t + \Theta s_t$，$0 \leqslant t \leqslant T$，$\{N_t,\ 0 \leqslant t \leqslant T\}$ 表示零均值、均方连续的高斯噪声过程；$\{s_t,\ 0 \leqslant t \leqslant T\}$ 表示已知的连续信号；Θ 为服从 $\mathcal{N}(\mu,\ \sigma^2)$ 分布的随机变量，且与噪声独立。

   (a) 假设检测问题

   $$H_0: Y_t = N_t, \qquad 0 \leqslant t \leqslant T$$
   $$H_1: Y_t = N_t + s_t, \quad 0 \leqslant t \leqslant T$$

   非奇异，已知 $\{Y_t,\ 0 \leqslant t \leqslant T\}$，求解 Θ 的 MMSE 估计。

   (b) 已知 λ 和 $\{\alpha_t;\ 0 \leqslant t \leqslant T\}$，且满足条件 $\int_0^T \alpha_t^2 \mathrm{d}t = 1$ 和 $\lambda > -\frac{N_0}{2}$，求解在自协方差满足下式时 Θ 的 MMSE 估计

   $$C_N(t,u) = \frac{N_0}{2}\delta(t-u) + \lambda\alpha_t\alpha_u, \quad 0 \leqslant t,u \leqslant T$$

   (c) 假设维纳过程 $\{N_t,\ 0 \leqslant t \leqslant T\}$ 的自协方差为 $C_N(t,\ u) = \frac{N_0}{2}\min\{t,\ u\}$，进一步假设 $s_t = -E_0\sin(\omega_0 t)/\omega_0 E_0 > 0$，对整数 n 有 $\omega_0 = n\pi/T$，讨论 σ^2 随 $E_0^2 T/N_0$ 变化时 Θ 的 MMSE 估计。

2. (a) 假设信号

   $$Y_t = A\sin(2\pi t + \Phi) + N_t, \quad 0 \leqslant t \leqslant 1$$

   其中 $\{N_t,\ 0 \leqslant t \leqslant 1\}$ 为零均值高斯随机过程且有连续自协方差函数 $C_N(t,\ u)$；随机变量 A 服从 $\mathcal{N}(\mu,\ \sigma^2)$ 分布且与 $\{N_t,\ 0 \leqslant t \leqslant 1\}$ 独立。假设 Φ 已知，求解已知 $\{Y_t,\ 0 \leqslant t \leqslant 1\}$ 下 A 的 MMSE 估计。

   (b) 用具有单位功率谱幅度的高斯噪声代替 $\{N_t,\ 0 \leqslant t \leqslant 1\}$ 并求解问题(a)。

   (c) 假设 Φ 是在 $[0,\ 2\pi]$ 均匀分布的随机变量，且独立于 A 和 $\{N_t,\ 0 \leqslant t \leqslant 1\}$，求解问题(b)。

3. 证明式(7.3.13)。

4. 利用式(7.3.20)求解积分方程(7.3.19)。

5. 假设随机过程 X_t 满足

   $$\frac{\mathrm{d}X_t}{\mathrm{d}t} = -aX_t + U_t, \quad t \geqslant 0$$

   其中 a 为一个正常数，X_0 是协方差为 $\sigma_0^2 > 0$ 的零均值随机变量；$\{U_t;\ t \geqslant 0\}$ 是一个与 X_0 独立的零均值白噪声且自协方差满足

   $$C_U(t,\tau) = \sigma_U^2\delta(t-\tau)$$

   假设观测模型如下

   $$Y_t = X_t + N_t, \quad t \geqslant 0$$

   其中 $\{N_t;\ t \geqslant 0\}$ 是零均值白噪声过程且独立于 X_0 和 $\{U_t;\ t \geqslant 0\}$，其自协方差为

   $$C_N(t,\tau) = \sigma_N^2\delta(t-\tau)$$

   (a) 解释由 Y_0^t 估计 X_t 时所采用的卡尔曼滤波的原理。

   (b) 用变量 a、σ_0^2、σ_U^2、σ_N^2 和时间 t 表达出均方估计误差和卡尔曼增益的表达式，画出 Kalman-Bucy 滤波器的原理框图。

   (c) 在稳态假设($t \to \infty$)下回答以上问题，分别考虑 $a^2 \gg (\sigma_U^2/\sigma_N^2)$、$a^2 = (\sigma_U^2/\sigma_N^2)$ 以及 $a^2 \ll (\sigma_U^2/\sigma_N^2)$ 的情形。

6. 考虑在习题 1(b)中 $\lambda = 0$ 的情况，并用 $\hat{\theta}_T$ 表示 Θ 的估计值随着 T 的增加($\hat{\theta}_T$ 可以利用 Kalman-Bucy 滤波器递归计算得到)。求稳态情况($T \to \infty$)下卡尔曼增益和 MMSE 估计结果。

7. 假设$\{S_t, t\geqslant 0\}$和$\{X_t, t\geqslant 0\}$是由下面微分方程得到的两个随机过程

$$\dot{S}_t = X_t, \quad t\geqslant 0$$

$$\dot{X}_t = 0, \quad t\geqslant 0$$

初始条件S_0和X_0是相互独立的随机变量，均服从分布$\mathcal{N}(\mu, \sigma^2)$。已知信号

$$Y_t = S_t + N_t, \quad t\geqslant 0$$

其中$\{N_t; t\geqslant 0\}$是功率谱幅度为$N_0/2$的零均值高斯白噪声，与S_0和X_0独立。

(a) 求解 Riccati 微分方程，得到下面的误差向量的协方差矩阵

$$\begin{pmatrix} S_t - \hat{S}_t \\ X_t - \hat{X}_t \end{pmatrix}$$

其中$\hat{S}_t = E\{S_t \mid Y_0^t\}$，$\hat{X}_t = E\{X_t \mid Y_0^t\}$。

(b) 假设关于S_0和X_0的初始信息较少，即在$\sigma^2\to\infty$的情况下估计$\hat{S}_t$和$\hat{X}_t$。

8. 已知信号模型如下

$$\dot{X}_t = -AX_t + U_t, \quad t\geqslant 0$$

$$Y_t = C_t X_t + N_t, \quad t\geqslant 0$$

其中$\{N_t; t\geqslant 0\}$和$\{U_t; t\geqslant 0\}$相互独立，分别是功率谱幅度为q和r的高斯白噪声；X_0是与$\{N_t; t\geqslant 0\}$和$\{U_t; t\geqslant 0\}$都独立的高斯过程。

(a) 假设对于所有t有$C_t = 1$，且A的取值为$a_1, a_2, \cdots, a_m$，对应的概率分别为$p_1, p_2, \cdots, p_m$。求指定$\{Y_s; 0\leqslant s\leqslant t\}$下$X_t$的稳态 MMSE 估计。

(b) 假设$C_t = \sin(\omega_0 t)$且A为正常数，求指定$\{Y_s; 0\leqslant s\leqslant t\}$下$X_t$的稳态 MMSE 估计(注意，这不是时不变滤波器)。

9. 用式(7.3.73)的滤波器解因果 Wiener-Hopf 方程(7.3.69)。

10. 证明式(7.4.44)。

参 考 文 献

Anderson, B. D. O., and J. B. Moore (1979), *Optimal Filtering* (Prentice-Hall: Englewood Cliffs, NJ).

Apostol, A. (1974), *Mathematical Analysis*, 2nd Ed. (Addison-Wesley: Reading, MA).

Ash, R. B., and M. F. Gardner (1975), *Topics in Stochastic Processes* (Academic: New York).

Barton, R. J., and H. V. Poor (1992), "On Generalized Signal-to-Noise Ratios in Quadratic Detection," *Mathematics of Control, Signals and Systems*, vol. 5, no. 1, pp. 81-91.

Beneš, V. E. (1981), "Exact Finite-Dimensional Filters for Certain Diffusions with Nonlinear Drift," *Stochastics*, vol. 5, pp. 65-92.

Beneš, V. E. (1987), "Nonlinear Filtering: Problems, Examples, Applications," Chapter 1 in *Advances in Statistical Signal Processing, Volume 1: Estimation*, H. V. Poor, Ed. (JAI Press: Greenwich, CT).

Bierman, G. J. (1977), *Factorization Methods for Discrete Sequential Estimation* (Academic: New York).

Billingsley, P. (1979), *Probability and Measure* (Wiley: New York).

Boekee, D. E., and J. C. Ruitenbeck (1981), "A Class of Lower Bounds on the Bayesian Probability of Error," *Inform. Sciences*, vol. 25, pp. 21-25.

Breiman, L. (1968), *Probability* (Addison-Wesley: Reading, MA).

Brockett, R., and J. M. C. Clark (1980), "The Geometry of the Conditional Density Equation," in *Analysis and Optimization of Stochastic Systems*, O. L. R. Jacobs, Ed. (Academic: New York).

Carlyle, J. W. (1968), "Nonparametric Methods in Detection Theory," Chapter 8 in *Communication Theory*, A. V. Balakrishnan, Ed. (McGraw-Hill: New York).

DeBruijn, N. G. (1961), *Asymptotic Methods in Analysis*, 2nd Ed. (North-Holland: Amsterdam).

Desoer, C. R. (1970), *Notes for a Second Course on Linear Systems* (D. Van Nostrand: Princeton, NJ).

Doob, J. L. (1953), *Stochastic Processes* (Wiley: New York).

Duncan, T. E. (1968), "Evaluation of Likelihood Functions," *Inform. Control*, vol. 13, pp. 62-74.

Duncan, T. E. (1970), "Likelihood Functions for Stochastic Signals in White Noise," *Inform. Control*, vol. 16, pp. 303-310.

Dunford, N., and J. T. Schwartz (1958), *Linear Operators, Part I* (Wiley: New York).

Feldman, J. (1958), "Equivalence and Perpendicularity of Gaussian Process," *Pacific J. Math.*, vol. 8, pp. 699-708.

Feldman, J. (1960), "Some Classes of Equivalent Gaussian Process on an Interval," *Pacific J. Math.*, vol. 10, pp. 1211-1220.

Ferguson, T. S. (1967), *Mathematical Statistics: A Decision Theoretic Approach* (Academic: New York).

Girsanov, J. V. (1960), "On Transforming a Certain Class of Stochastic Process by Absolutely Continuous Substitution of Measures," *Theory Prob. Appl.*, vol. 5, pp. 285-301.

Goodwin G. C., and K. S. Sin (1984), *Adaptive Filtering, Prediction and Control* (Prentice-Hall: Englewood Cliffs, NJ).

Grenander, U. (1981), *Abstract Inference* (Wiley: New York).

Hajek, J. (1958), "On a Property of Normal Distributions of Any Stochastic Process," *Czech. Math. J.*, vol. 8, pp. 610-617.

Hajek, J., and Z. Sidak (1967), *Theory of Rank Tests* (Academic: New York).

Hampel, F. R., *et al.* (1986), *Robust Statistics: The Approach Based on Influence Functions* (Wiley: New York).

Hazewinkel, M., and S. Marcus (1982), "On Lie Algebras and Finite-Dimensional Filtering," *Stochastics*, vol. 5, pp. 29-62.

Helstrom, C. W. (1968), *Statistical Theory of Signal Detection*, 2nd Ed. (Pergamon: Oxford, UK).

Honig, M. L., and D. G. Messerschmidt (1984), *Adaptive Filters: Structures, Algorithms, and Applications* (Kluwer: Boston).

Huber, P. J. (1965), "A Robust Version of the Probability Ratio Test," *Ann. Math. Stat.*, vol. 36, pp. 1753-1758.

Huber, P. J. (1981), *Robust Statistics* (Wiley: New York).

Huber, P. J., and V. Strassen (1973), "Minimax Tests and the Neyman-Pearson Lemma for Capacities," *Ann. Statist.*, vol. 1, pp. 251-263.

Kailath, T. (1966), "Some Integral Equations with 'Nonrational' Kernels," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 442-447.

Kailath, T. (1969), "A General Likelihood Ratio Formula for Random Signals in Gaussian Noise," *IEEE Trans. Inform. Theory.*, vol. IT-15, pp. 350-361.

Kailath, T. (1971), "The Structure of Radon Nikodym Derivatives with Respect to Wiener and Related Measures," *Ann. Math. Stat.*, vol. 42, pp. 1054-1067.

Kailath, T. (1981), *Lectures on Wiener and Kalman Filtering* (Springer-Verlag: New York).

Kassam, S. A., and H. V. Poor (1985), "Robust Techniques for Signal Processing: A Survey," *Proc. IEEE*, vol. 73, pp. 433-481.

Kassam, S. A., and J. B. Thomas (1980), *Nonparametric Detection: Theory and Applications* (Dowden, Hutchinson & Ross: Stroudsburg, PA).

Kendall, M. G. (1948), *Rank Correlation Methods* (Griffin: London).

Kendall, M. G., and A. Stuart (1961), *The Advanced Theory of Statistics, Vol. 2* (Hafner: New York).

Kobayashi, H., and J. B. Thomas (1967), "Distance Measures and Related Criteria," *Proc. 5th Ann. Allerton Conf. Circuit and System Theory*, Monticello, IL, pp. 491-500.

Kullback, S. (1959), *Information Theory and Statistics* (Wiley: New York).

Lehmann, E. L. (1983), *Theory of Point Estimation* (Wiley: New York).

Lehmann, E. L. (1986), *Testing Statistical Hypotheses* (Wiley: New York).

Lipster, R. S., and A. N. Shiryayev (1977), *Statistics of Random Processes*

I: General Theory (Springer-Verlag: New York).

Ljung, L., and T. Soderstrom (1982), *Theory and Practice of Recursive Identification* (MIT Press: Cambridge, MA).

Lovitt, W. V. (1950), *Linear Integral Equations* (Dover: New York).

Lugannani, R., and S. Rice (1980), "Saddle Point Approximation for the Distribution of the Sum of Independent Random Variables," *Adv. Appl. Prob.*, vol. 12, pp. 475-490.

Lukacs, E. (1960), *Characteristic Functions* (Hafner: New York).

Marcus, S. I. (1984), "Algebraic and Geometric Methods in Nonlinear Filtering," *SIAM J. Control Optimization*, vol. 22, pp. 817-844.

Martin, R. D., and S. C. Schwartz (1971), "Robust Detection of a Known Signal in Nearly Gaussian Noise," *IEEE Trans. Inform. Theory,* vol. IT-17, pp. 50-56.

Mazo, J. E., and J. Salz (1965), "Probability of Error for Quadratic Detectors," *Bell Syst. Tech. J.,* vol. 44, pp. 2165-2186.

Nevel'son, M. B., and R. Z. Has'minskii (1973), *Stochastic Approximation and Recursive Estimation* (American Mathematical Society: Providence, RI).

Noether, G. E. (1955), "On a Theorem of Pitman," *Ann. Math. Stat.,* vol. 26, pp. 64-68.

Oppenheim, A. V., and R. W. Schafer (1975), *Digital Signal Processing* (Prentice-Hall: Englewood Cliffs, NJ).

Papoulis, A. (1986), *Probability, Random Variables and Stochastic Processes* (McGraw-Hill: New York).

Parzen, E. (1962), *Stochastic Processes* (Holden-Day: San Francisco).

Schweppe, F. C. (1965), "Evaluation of Likelihood Functions for Gaussian Signals," *IEEE Trans. Inform. Theory,* vol. IT-11, pp. 61-70.

Shepp, L. A. (1966), "Radon-Nikodym Derivatives of Gaussian Measures," *Ann. Math. Stat.,* vol. 37, pp. 321-353.

Skorohod, A. V. (1974), *Integration in Hilbert Space* (Springer-Verlag: New York).

Slepian, D. (1958), "Some Comments on the Detection of Gaussian Signals in Gaussian Noise," *IRE Trans. Inform. Theory,* vol. IT-4, pp. 65-68.

Stranovich, R. L., and Yu. G. Sosulin (1965), "Optimal Detection of a Diffusion Process in White Noise," *Radio Eng. Electron. Phys.,* vol. 10, pp. 704-713.

Tantaratana, S. (1986), "Sequential Detection of a Positive Signal," Chapter 7 in *Communications and Networks: A Survey of Recent Advances,* I. F. Blake and H. V. Poor, Eds. (Springer-Verlag: New York).

Thomas, J. B. (1971), *An Introduction to Applied Probability and Random Processes* (Wiley: New York).

Thomas, J. B. (1986), *An Introduction to Applied Probability* (Springer-Verlag: New York).

Trench, W. F. (1964), "An Algorithm for the Inversion of Finite Toeplitz Matrices," *J. SIAM,* vol. 12, pp. 515-522.

Van Trees, H. L. (1968), *Detection, Estimation and Modulation Theory, Part I* (Wiley: New York).

Viterbi, A. (1968), *Principles of Coherent Communications* (Wiley: New York).

Wong, E. (1983), *Introduction to Random Processes* (Springer-Verlag: New York).

Wong, E., and B. Hajek (1985), *Stochastic Processes in Engineering Systems* (Springer-Verlag: New York).

Yao, K., and R. M. Tobin (1976), "Moment Space Upper and Lower Bounds for Digital Systems with Intersymbol Interference," *IEEE Trans. Inform. Theory,* vol. IT-22, pp. 65-74.